IN SUPPORT OF GARCH FOR VOLATILITY ANALYSIS AND FORECASTING

Tushar Saini[1]
Final Year BCOM F&A Student
Department of Professional Studies
Christ University, Bangalore, India

Thangjam Ravichandra[2]
Assistant Professor
Department of Professional Studies
Christ University, Bangalore, India

ABSTRACT

The study projects the misuse of linear regression for volatility analysis and forecasting. On testing the fundamental assumptions of linear regression in time series,Linear regression is not deemed fit for the use because of the violation of the modelprinciples. The Generalized Autoregressive Conditionally Heteroskedasticity (GARCH 1,1) model is considered for estimating volatility. GARCH (1,1) has proved to be an ideal model in the GARCH family as its more consistent and efficient with its results. A relationship has been derived between the NIFTY AND NIFTY F&O market movements. Therefore, spearman's correlation is performed on the prices of spot market index (NIFTY) and the volatility of the derivative market index (NIFTY F/O). Reasonablejustification for using spearman's correlation is because the variance data of the NIFTY F/O represents heavy positive skewness, which violated the fundamental assumption of the Pearson's correlation. The result of the correlation was significant. The tools used for the work are:linear regression, Durbin Watson test for autocorrelation, scatter plots, outliers labeling, normal distribution plots, regression plot, graphical representation and GARCH (1,1) model. As a result a composite model has been framed with sound justification for estimating the volatility and movement impact of both the indexes.

Index Words: Linear Regression, Volatility Analysis, GARCH, NIFTY, NIFTY F&O, Spot market index, Correlation

I. INTRODUCTION

Volatility in financial markets is amongst one of the most important element to be considered regarding the decision made for allocations of funds. Unfortunately, the models used to estimate volatility in the financial markets are often wrongly allocated. The confusion lies in the use of regression for estimating the volatility in the time series data having derivative relationship. The understanding of the nature of the data sets used and their relationship is important before expressing a model to estimate its volatility.

Volatility can be defined as a statistical measure of deviation of returns for a given variable. The volatility is the key indicator of the fluctuation of the price of stocks or the market index. Financial markets always represent the element of risk. Without risk, financial markets loose their capability of impressing the investors. To comprehend it in practice one must understand the measure of risk. Which is very evidently shown by the standard deviation or the variance of the returns. Most of the investors and the traders are interested in the prices of the financial assets. The price nonetheless is a function of the variance of the security. Volatility is the key to the price because it depicts the data generation process of a security and moreover also helps the investors to model their leverage. With the introduction of derivative trading came transparency in the system. Regulations have increased the participation in the capital markets, leading to more financial deepening and widening.On the other hand it has also made it very complicated over time with the involvement of high-end mathematics, physics, statistical models and functions. Such complications leading to huge disasters too, a classic example being Long Term Capital Management (LTCM), a Hedge fund with an investment strategy of Convergence Arbitrage. The company was shut down with an estimated loss of about $1 billion. Other examples such as the Recession of 2008, which occurred due to the subprime mortgages leading to a bloodbath in the U.S stock markets, the 1929 stock market crash, the 1940 German economic crisis, the China stock market crash etc. Therefore many ground breaking concepts were derived in this field of capital investments. To merge the gap between uncertainty followed by irrationalhuman behavior of the investors towards creating the imbalance in the markets, a strong conceptual framework accepted to be served by academics which needed to be balanced using fundamentally strong models to arrive at conclusions which would enhance the investors capability to invest and improve their confidence. Amongst those few concepts lie the strength of the GARCH model and regression, that studies volatility as the key measure over the judgments, principles and fundamentals of mathematic and statistics. Engle's (1982) study on ARCH and Bollerslev's (1986) study on GARCH models concluded that both the models are meant to deal with the non-stationary data as found in the real world scenario in financial markets. He further pointed out that these models have become widespread tools for dealing with time series Heteroscedasticity which is the circumstance in which the variability of a variable is not equal across values of a second variable that forecasts it. The ARCH and GARCH models treat Heteroscedasticity as a variance to be modeled. The goal of such models is to provide a measure for volatility like standard deviation, beta etc, that can be used extensively in financial decisions concerning analysis of risk, selection of portfolio and pricing of derivative in real life scenarios with more dependency and confidence. After ascertaining the volatility, the use of the information to make strategic policies and further investment related choices; the concept of correlation stands fit to be used as it represents the strength of the relationship between two data sets. Therefore estimating the volatility of NIFTY F/O and correlating it with the price movement of NIFTY provides a complete model of volatility and testing the strength of the time series to make sound decisions, and also to forecast the future movement of the time series. Although, in this work, the research has been restricted to the use of empirical data for estimating variance and testing the strength of movement of NIFTY F/O variance and NIFTY INDEX points.

II. REVIEW OF LITERATURE

The role of capital markets has influenced the economy in aggressive ways. Today, though in India not more than 2% of the total population invest their funds in capital markets, the progress of this specific niche has been incredible. The index futures were introduced in the year 2000-01 with 90,580 contracts being traded and by the end of 2011-12 the contracts that were being traded in the market was 13,57,95,014 (option future and other derivatives, hull and Basu). Such empirical data reflect the potential of the capital markets in the country. The same trend is seen in the transaction and turnover of the spot market. As the role of the Capital market is very imperative to sustain economy and so are the methods used for making sound decisions, the issues presented by linear regression can majorly affect the decision making of the investor. The problems with linear regression is discussed in the work issued by R.H. Shumway and D.S. Stoffer, *Time Series Analysis and Its Applications*; Springer Texts in Statistics (2011). The assumptions of the linear regression are tested against a study conducted by Michael A. Poole and Patrick N. O'farrell, *The Assumption of Linear Regression Model* (1970). The paper presents the assumptions of the Regression Model at a comprehensive scale, and stated that even if a single assumption of the Regression model is violated then the model shall necessarily be rejected. But in the real world, the model is accepted as usually at least one assumption is violated. Strictly speaking, to state a model to be unbiased in nature, all the assumptions must be met as even a single unmatched assumption can lead to distraught results in estimation and forecasting. The reason why time series for forecasting was introduced was because of the fundamental problem in dated series that the residuals usually provided auto-correlated feature, which technically reduces the size of the number of observations hence reducing the sample of the work. The paper presented by *Hing and stein* (1999) stated that mostly the returns of the stocks are positively correlated, and using this research as extended to my study, testing for autocorrelation actually provided that the Residuals of both the NIFTY F/O index and NIFY Index is positive Autocorrelation, which usually presents a problem of over estimation of the forecasted values. *Thomas and Bernard* (1990) , argued that the presence of autocorrelation can complicate the results of the research. Therefore looking at all such problems with autocorrelation,Jennifer Neville, Ozgur Simsek, and David Jensen in their issue of *Autocorrelation and Relational Learning: 'Challenges and Opportunities presented the actual problem with autocorrelation' argued about the presence of autocorrelation in time series*argued that Autocorrelation violates the fundamental principle of independence, which is a base assumption for most of the classical models including linear regression. In the paper presented by P. Reinhard Hansen and A. Lunde titled*Does anything beat a GARCH (1,1)? (*2001), a comparative study on all the time series forecasting models, i.e. the whole GARCH family, they concluded that GARCH (1,1) is the most efficient model when compared on the basis of efficiency and forecasting. Therefore on the basis of the findings of the result in the above-mentioned study, I used the standard GARCH (1,1) model. Using the reference of the paper *parameter estimation of GARCH model* by YiyangYang (2012) the criterion of the model were set as the trial values of omega, alpha and beta.Calculation of likelihood estimator and the results of the estimation, to arrive at the volatility figures. The paper presented a sequential way to estimate parameter of non-stationary time series. *Campbell Lo, and MacKinlay* (1997, p.481), argued that it is statically and mathematically incorrect to measure the performance of a variable using conditional volatility in capital markets where the markets more or less functions independent of the economic system, with immense structural breaks and noise in trading. The *ACD (Autoregressive Conditional Duration) model* of Russell and Engle (1998) was developed to describe relationships in the durations between random occurring events. For high frequency and financial data this research has become a benchmark paper for estimation and preparation of portfolio and investment strategies. Sibani and Uma (2007) used GARCH and OLS techniques to record the time-varying nature of volatility and the phenomenon of volatility clustering of the Nifty Index, which was compared with the establishment of derivative indexes in India. The results represented that there is no significant change in the time varying variance of the spot market of the Nifty Index, but the structure of volatility has changed to a minor level. The study also reported that the impact of new information in the market is very quick compared to the empirical era. Mallikarjunappa and Afsal (2007) examined the volatility behavior of the Indian market by focusing on the CNX IT Index, which is a sectorial index, and found that underlying volatility increases with the onset of futures trading. Even though the result of the studies have shown different results over the same indexes, if we consider the other factors such as time sampling, the methods used under the GARCH models, the error rates, hypothesis testing etc. The changes or variations in the results are bound to be seen. The work presented by J*an Hauke, Tomasz Kossowski*(2011), was used as the benchmark for selection of a correlation technique between Spearmen and Pearson correlation.

III. STATEMENT OF THE PROBLEM

Without volatility the market freezes, as there would be no sentiments in the participants for making gains over the trading. Also there is no universally accepted phenomenon for this cause, the most reliable explanation being the changes in the demand and supply. As a matter of fact, with economic and statistical data, varying volatility is a more common phenomenon than constant volatility, and accurate modeling of time-varying volatility is of great importance than constant volatility in financial engineering. The models such as Black Scholes are based on the assumptions of constant volatility i.e. even though the price of the underlying asset might change over time its volatility will remain constant. Such assumption makes the model limited of its use mostly just till the academics. When considering the time series specifically the financial market data provides some drastic errors. Such errors can be a disaster for the person relying on them therefore, the statement of problem being;which is the best method for estimating the time series values and variances?Deriving a method to rely on the relationship between the volatility of NIFTY F/O and the price movement of NIFTY index. The reason behind using estimated volatility of NIFTY F/O to establish a link between the spot and derivate market index movements and fluctuation is because the volatility represents the data generation process of a time series. If a link between the volatility and price movement of two different time series is estimated, then the conclusion can be derived both ways and this helps in making estimations and forecasts for not only one time series but also for the other. Ex. if the volatility estimation of NIFTY F/O increases, the price of NIFTY decreases. Then this information can also read in a way that as the price of NIFTY decreases the volatility of NIFTY F/O increases. Therefore it helps to understand the relationship between both the time series well.

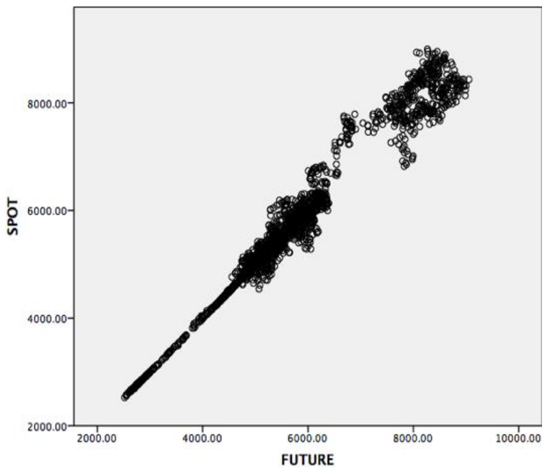
IV. OBJECTIVES OF THE STUDY

1) Testing the assumption of linear regression in time series data.
2) Estimating volatility using GARCH model.
3) Analyzing which model is a good fit for estimating volatility: Linear regression or GRACH?
4) Using Spearman's correlation to express the relationship between NIFTY price movements and NIFTY F/O volatility.

V. ANALYSIS

A linear regression attempts to model the relationship between the dependent variable and an independent variable using mathematical modeling. The established relationship between two variables in linear regression does not necessarily mean that changes in one variable will impact the other variable. Rather it expresses a significant relationship between the two variables, which could be expressed mathematically. A prerequisite of the linear regression is that the series must be a in a linear mode, which can be explained with a statistical tool called scatterplot fig1.1

Scatterplot, fig1.1



Simpler linear regression can be conducted with only 2 variables. However there is also an extension to this method called multiple linear regressions where more than 2 variables can be adjusted in the model. Sir Francis Galton pioneered the concept. Linear regression is used in many areas such as analyzing the trend and projecting sales, risk assessment, analyzing the impact of price change and so on. The reason why Simple Linear regression is used is to analyze whether the movement of the data set reflects linear relationship, finding the residuals/errors rates of the data and the type of relationship data reflects. Result of The linear regression is presented in the fig 1.2 and fig 1.3

EQUATION TABLE, fig 1.2

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	111.906	23.980		4.667	.000	64.878	158.934
	FUTURE	.984	.004	.983	245.199	.000	.976	.992

a. Dependent Variable: SPOT

EQUATION TABLE, fig 1.3

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.983[a]	.967	.967	263.84132	.967	60122.588	1	2038	.000

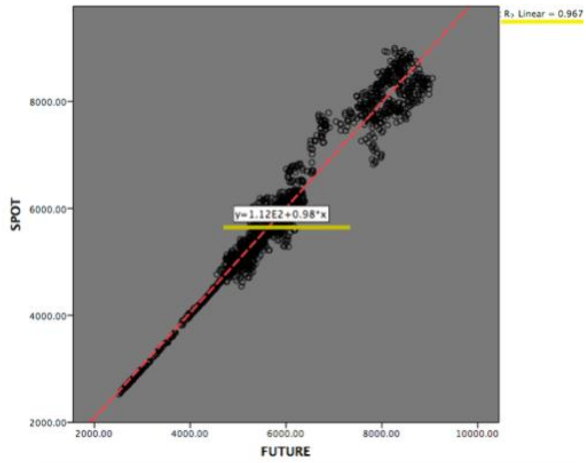
a. Predictors: (Constant), FUTURE

The formula for simple linear regression

$$Y = \beta 0 + \beta 1 X + e$$

Where the Dependent variable is also called as response variable, criterion, or label and is denoted by 'Y'. The independent variable is also referred to as covariates. 'X' represents it. Where 'e' is the error while calculating 'Y'. Therefore the equation generated for linear regression is presented in fig 1.4.

REGRESSION LINE, fig 1.4



R square in the fig1.4, represents coefficient of determination meaning it represents the average of SSR (Sum of regression) over SST(total sum of square). It represents percentage of sum of squares that can be depicted using the regression equation. In the data presented, the R square value means that 96.7% of the total sum of squares can be explained using this particular regression model and the remaining 3.3% could not be captured by the analysis hence it is the error rate.

Overall concluding; the results of the linear regression model seems to provide reasonable results. .984 represents Beta, which means that change of 1 point in the future index would lead to a change of .984 points in the spot index. The intercept depicts a high value of 111.906, which represents that on zero changes in the future index, spot index would have a variation of 111.906 points. Practically, the intercept may or may not be required to fit the real world requirements.

After generating an equation for linear regression, which seems fine to the untrained eye, the analysis can be a deception in real practice. Therefore I cross-examined whether the model is fit to be used. For the purpose of this we check the model against the standard assumption of the linear regression technique.

NOTE: If any of these assumptions are violated then the model is not fit for the further analysis to be concluded.

TESTING THE ASSUMPTIONS OF LINEAR REGRESSION FOR THE ABOVE PROJECTED ANALYSIS:

1) The variables used shall be measured at a continuous level: Meaning the unit used to represent the variable shall be continuous in nature .i.e. the units shall not change. If in a series, for example days are used to measure the data then the series should continue with days only. Ensuring the units should not change to the other modes of measurement.For the purpose of this study the unit used is days closing price, which is maintained through out the time series from the first to the last observation. *Therefore there is no violation of this assumption.*

2) There needs to be a linear relationship between the variables: The linear relationship between the variables can be estimated using a tool called scatterplot by plotting the dependent variable against an independent variable*(fig 1.1)*. The analysis projected for the data series taken into consideration represents the positive linear relationship between the data. *Therefore there is no violation of this assumption.*

3) Significant outliers should not be present: An outlier is a data point in the plot representing a value, which differs significantly from the values predicted by the regression equation. The outliers may have a negative effect on the results of the regression i.e. making the regression analysis inaccurate and over or under estimated. The method used for detecting outliers is *outliers labeling by Turkey J.W*, *explanatory data analysis*. Using a multiplier or g value to be 1.5 the analysis proved that there are significant outliers present, *therefore breaching the assumptions of linear regression.*

4) The observations should be independent of each other: in our case the time series shall be independent i.e. there should not exist any autocorrelation in each of the series errors, although the presence of autocorrelation in time series is usually seen as a natural phenomenon. Autocorrelation represents the dependency of one value on the value of the same time series. Just like auto regression, autocorrelation is used to represent the correlation between the same values of a single time series. The criterion used for testing the presence of autocorrelation in a regression model *is Durbin Watson Test*. The values of the test range from 0 to 4. The assumption followed by Durbin Watson test is that errors present in the regression equation are generated by first order autoregressive process and are presented at equally spaced time periods, that is; $\varepsilon t = \rho \varepsilon t-1 + at$ where εt represents the error term in the model at time period t, 'at' is an NID (0, σa2) random variable, and ρ (|ρ| < 1) represents the autocorrelation parameter.

DURBIN WATSON TEST, fig1.5

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.983[a]	.967	.967	263.76616	.104

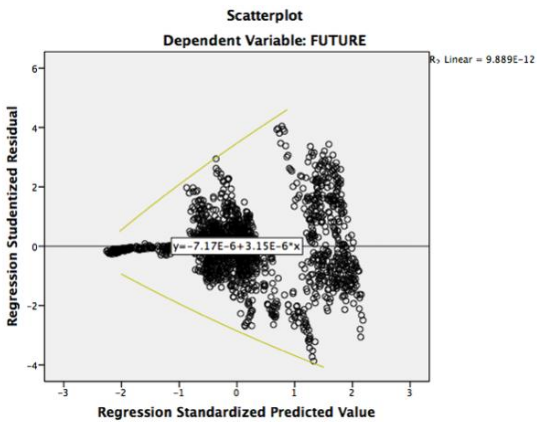
a. Predictors: (Constant), SPOT

b. Dependent Variable: FUTURE

As stated, the fig 1.5 Durbin Watson test presented the value of .104, stating high positive autocorrelation in the time series. *Therefore it violates the underlying assumption of the regression model*. This significant value of autocorrelation may provide absurd results in the forecast, also over estimating the result.

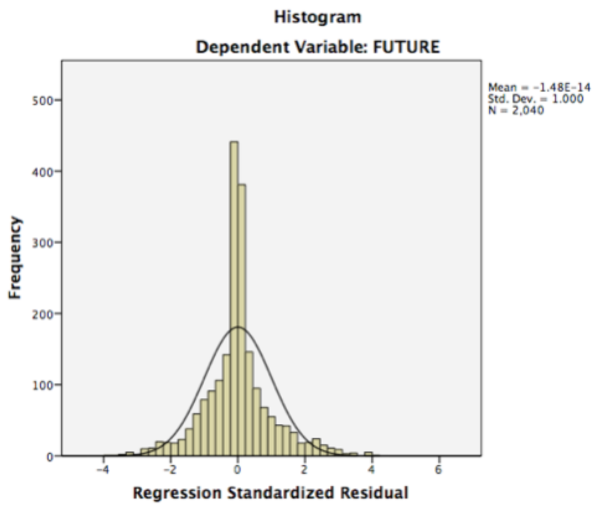
5) **Data needs to show homoscedasticity: h**omoscedasticity is a statistical term depicting that the error variances are similar over the samples. This type of error structure is most often assumed in conventional methods for modeling relationship between two variables, but the assumption does not always stand correct. As presented below the pattern formed by the residuals shows clustering initially followed by dispersion**.** Here the errors represent Heteroscedasticity and not Homoscedasticity as presented below in fig 1.6, *therefore violating the assumptions of linear regression fundamentals*.

HETEROSCEDASTICITY, fig 1.6



6**) Check that the residuals of the regression line are nearly normally distributed:** using the histogram plot, as shown below in fig 1.7, for testing this assumption. Although thereis slight *deviations it is a considerable fit for the model.*

HISTOGRAM, fig 1.7



VI. CONCLUSION TO REGRESSION MODEL

The final conclusion to the model is that there are various models that usually use regression for forecasting the volatility, but technically regression would provide non-satisfactory results, the reason being breach of assumptions of regression model. When tested against the fundamentals of regression this modeling might provide absurd results, which surely cannot be relied upon when it comes to the investment decision process. As represented above, assumptions such as; there should not be any outliers, observations should be independent of each other, and data needs to show homoscedasticity was not fulfilled.

Therefore this is the reason why regression should not be used with non-stationary series.

ESTIMATING AND FORCASTING VOLATILITY USING GARCH(1,1) MODEL

GARCH time series models are becoming widely used in econometrics and finance because they consider randomly changing volatility. The frequency distribution for GARCH model can be daily, weekly, monthly and even yearly. GARCH represents Generalized Autoregressive Conditionally Heteroskedasticity. In general, the term 'Heteroskedasticity' can be referred to as time-varying variance or in other words volatility. The term 'Conditional' represents the dependence on the events of the past, and 'Autoregressive' describes a feedback system that incorporates past observations into the present. A common phenomenon in time series is volatility clustering, in which there are periods of high volatility followed by high volatility and low volatility followed by low volatility. The aim of a GARCH model is to encompass this volatility clustering .The proof is presented in the details regarding the 4th movement of the model. Along with providing estimations of volatility generation, the model also assigns weights that exponentially decline i.e. in descending order. Therefore most recent shock has more impact on the model.

In estimating the parameters of the GARCH (1,1) model the maximum likelihood estimation was implemented. For any given day, the returns for the next day is assumed to be conditionally distributed as a normal variable with mean zero and a variance rate. The reason GARCH process is used to model volatility can be addressed by viewing the strengths of the model and how this correlates with the behaviors, assumptions, and characteristics of a time series of financial data. The ARCH/GARCH family of modeling techniques is one of the most popular and fundamental modeling approaches for volatility forecasting and estimation. From the accurate estimation of future volatility in financial assets, one can have effective hedging against risk, portfolio optimization, and pricing options. Since the unconditional variance of returns is $E[\sigma 2] = \alpha 0/(1 - \alpha 1 - \beta 1)$, we can write the GARCH (1,1) equation yet another way; $\sigma 2 = w + \alpha a2 + \beta \sigma 2$, where $\alpha 0 > 0, \alpha 1 > 0, \beta 1 > 0$, and $\alpha 1 + \beta 1 < 1$, so that our consecutive period forecast of variance is a combination of our last period forecast and last period's squared return. For variance the relationship can be stated as $Var(at) = \alpha 0/(1 - \alpha 1 - \beta 1)$. Since the ln L function is a monotonically increasing function of L, we can maximize the log of the likelihood function and for a GARCH (1,1), we can substitute σ2 = α +α a2 +β σ2 into the above equation, and the likelihood function is only a function of the returns, rt and the parameters, as presented in fig 1.8

PARAMETERS OF GARCH, fig 1.8

PERIMETER	VALUES
w	0.0000024777
beta	0.912753
alp	0.074766
long run variance per day	0.00019851
volatility per day	0.014089

REPRESENTING ESTIMATED VOLATILITY AND NIFTY INDEX MOVEMENT GRAPHICALLY

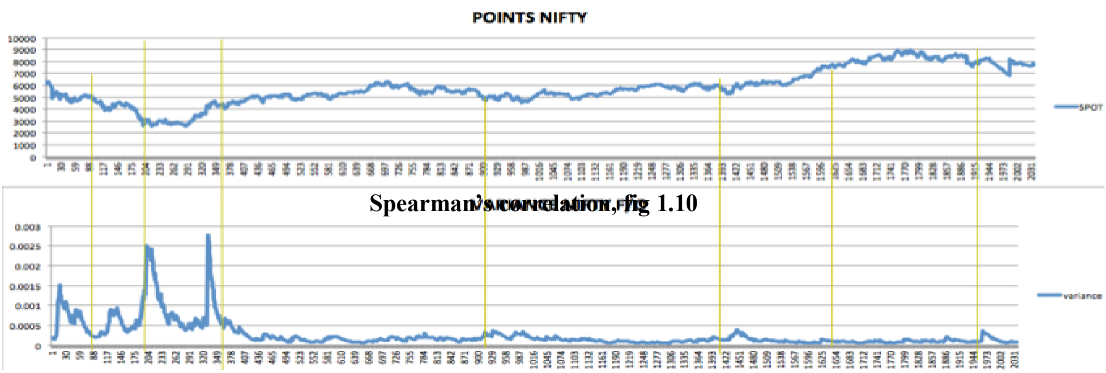
As presented in the above charts it can be observed that as and when the variance of the NIFTY F/O increases, the movement of NIFTY shows an inverse relation (fig 1.09). From this it can be estimated that the variables, NIFTY SPOT price movement and volatility of NIFTY F/O have an inverse relation. Testing this would provide a valid argument. Therefore we perform spearman's correlation (fig 1.10). The reason for using spearman's correlations is because the variance data of the NIFTY F/O represents heavy positive skewness, which violated the fundamental assumption of the Pearson's correlation. Therefore we conduct the substitute correlation called spearman's correlation to test the results. To test the efficiency of the results we shall consider the p value to prove statistically whether the values presented are significant, corrected or not.

Volatility and price moment of NIFTY &NIFTY F/O, fig 1.09

Correlations

			SPOTINDEX	VARIANCEFUTURE
Spearman's rho	SPOTINDEX	Correlation Coefficient	1.000	-.719**
		Sig. (2-tailed)	.	.000
		N	2038	2038
	VARIANCEFUTURE	Correlation Coefficient	-.719**	1.000
		Sig. (2-tailed)	.000	.
		N	2038	2038

**. Correlation is significant at the 0.01 level (2-tailed).



Spearman's correlation, fig 1.10

INTERPRETING THE RESULTS OF SPEARMAN'S CORRELATION

As presented in the above table the correlation presents a strong negative correlation of -.719 with a significance level of .000. Therefore we reject null hypothesis and accept alternate hypothesis, which states that there is significant correlation between the variance of derivative index and the price of spot index.

CONCLUSION

Linear regression can be used in time series modeling, but only after some major adjustments are made in the criteria where its fundamental assumptions are violated. Such adjustments complicate the model. Therefore using linear regression for time series modeling is not considered to be a statistically correct method for forecasting or analyzing the variability of time series, as adjustments make the model less trustable. The use of GARCH(1,1) process is considered to be a well-accepted idea. GARCH(1,1) model provides more practical results when compared to linear regression. GARCH(1,1) is the most trustable model in the real word. Although there are other better models available in the GARCH family, GARCH(1,1) seems to be the best fit with respect to efficiency of the results. Considering the spearman correlation for the volatility and price as 2 different data sets, it provides the relation between the 2 time series, in our case, spot and derivative market.

The research proves that there is a significant relationship between spot market price movement and that of volatility of the future market, which could also be seen as a feature of the trend proposed in the plot above. Overall the use of GARCH(1,1) model with spearman correlation proves to be a successful combination for estimating volatility and determining its impact on the other time series in use, which in our case is the spot market index. The research provided successful and significant results.

VII. REFRENCES

1. http://www.nseindia.com/products/content/equities/indices/historical_index_data.html
2. http://www.livemint.com/Money/VBSd0BJgg8KBBhy9VMlmRO/Top-10-biggest-falls-in-Sensex-history.html
3. http://pages.stern.nyu.edu/~rengle/vol_paper_30jan01.PDF |
4. http://faculty.tuck.dartmouth.edu/images/uploads/faculty/jonathan-lewellen/Momentum.pdf
5. http://eclectic.ss.uci.edu/~drwhite/courses/Regression_and_Autocorrelation.pdf |
6. https://www.cs.purdue.edu/homes/neville/papers/neville-et-al-srl2004.pdf
7. http://carecon.org.uk/UWEcourse/Problems.pdf|
8. http://www.bauer.uh.edu/rsusmel/phd/HansenLunde_Garch.pdf |
9. https://www3.nd.edu/~rwilliam/stats2/l26.pdf |
10. http://web.nchu.edu.tw/~finmyc/tim_GARCH.pdf|
11. www.telemed.no/getfile.php/917456.357.../Appendix-Definitions.pdf
12. http://www.cluteinstitute.com/ojs/index.php/IBER/article/viewFile/7994/8048| 13. http://www.bauer.uh.edu/rsusmel/phd/garch101 (ENGLE).pdf