EFFICIENCY OF INDIAN STOCK MARKETS: A STUDY OF NSE

INTRODUCTION:
The efficient market hypothesis states that asset prices in financial markets should reflect all available information; as a consequence, prices should always be consistent with ‘fundamentals’. The efficiency of stock market in economic development cannot be overemphasized. Efficient Stock Markets provide the vehicle for mobilizing savings and investment resources for developmental purposes. They afford opportunities to investors to diversify their portfolios across a variety of assets. This has the potential to reduce the cost of capital through lower risk premiums demanded by supplier of capital. In general, ideal market is the one in which prices provide accurate signals for resource allocation so that firms can make productive investment decision and investors can choose among the securities under the assumption that securities prices at any time fully reflect all available information. A market in which prices fully reflect all available information is called efficient.

The proposed study intends to investigate whether prices in National Stock Exchange i.e. S & P CNX Nifty, its constituents and other indices of NSE follow a random walk as required by the market efficiency. It will compare the results with NYSE and Chinese Stock Market especially Shanghai Stock Exchange being the oldest stock exchange in China, to get additional understanding of the market efficiency. If the null hypothesis of random walk is rejected, linear and non-linear modeling of the serial dependences will be conducted using ARIMA and GARCH models. Forecasts based on the best fitting models will also be compared for accuracy.

LITERATURE REVIEW
Fama (1965) propounded his famous efficient market hypothesis for US securities, a number of empirical research have been carried out to test its validity, mainly in the developed countries with booming financial markets (Summers, 1986; Fama and French, 1988; Lo and Mackinlay, 1988). Fama classified stock market efficiency into three forms. They are namely ‘weak form’, ‘semi-strong form’ and ‘strong form’. The classification depends upon the underlying assumptions relating to information set available to market participants. Each information set here is more comprehensive than the previous one.
Weak Form Efficient Market Hypothesis, which is also known as Random Walk Hypothesis (RWH) states that present prices of securities fully reflect information contained in their historical price. Therefore, the best predictor of the future price is the present price. It is not possible for the investors to design profitable strategy on the basis of past price of a security. Random walk is the path of a variable over time that exhibits no predictable pattern at all. If a price moves in a random walk, the value of price in any period will be equal to the value of price in the period before, plus or minus some random shocks. Semi-strong Efficient Market Hypothesis claims that prices of securities incorporate publicly available information, while strong-form holds that all the information set whether public or/and private are enveloped in the market prices of securities. Hence the prediction of future price conditional on past information is not advantageous to market participants. The more efficient capital market is more random which makes the market return more unpredictable. In the most efficient stock market, future prices will be purely random and the price formation can be assumed to be a stochastic process with mean in price change would be equal to zero.

Fama (1991) renamed the market efficiency studies into three categories. The first category involves the tests of return predictability; the second group contains event studies and the third tests for private information.

The concept of random walk was first developed by Bachelier (1900). He found that a successive price change between two periods is independent with zero mean and variance depends upon interval between two periods. The early studies on testing the weak form efficiency on the developed stock markets, generally agree with the support of weak-efficiency of the market considering a low degree of serial correlation (Cootner, 1962; Fama, 1965 and 1970). Porterba and Summers (1988) confirmed the presence of mean reverting tendency and absence of random walk in the U.S. Stocks. Lo and McKinney (1988) proposed variance ratio test to test random walk hypothesis. Their findings provided the evidence against random walk hypothesis for the entire sample period of 1962 to 1985. Fama and French (1988) discovered that forty percentage of variation of longer holding period returns were predictable from the information on past returns for U.S. Stock markets. Cambel (1991) used variance decomposition method for stock return and concluded that the expected return changes in persistent fashion.

Kim, Nelson and Startz (1991) examined the random walk pattern of stock prices by using weekly and monthly returns in five Pacific-Basin Stock Markets. They found that all stock markets except Japanese stock market did not follow random walk. Pope (1989) noted that the traditional tests of random walk model such as serial correlation and run test are
susceptible to error because of spurious autocorrelation induced by non-synchronous trading. Shiller and Perron (1985) and Summer (1986) have shown that such tests have relatively little power against interesting alternative hypothesis of market efficiency. Cutler, Porterba and Summer (1990) found evidence of the mean reversion and predictability of the US stock market return. David Walsh (1997) employed variance ratio test to test the null hypothesis of random walk in the Australian Stock Exchange covering various sampling intervals and data period during January 1980 to December 1995. His result suggested that many indices of the stock exchange returned to random walk during October Crash 1987.

Madhusudan (1998) found that BSE sensitivity and national indices did not follow random walk. Using correlation analysis on monthly stock returns data over the period January 1981 to December 1992, Olowe (1999) shows that the Nigerian stock market is weak form efficient. Bhanu Pant and Dr. T.R.Bishnoy (2001) analyzed the behaviour of the daily and weekly returns of five Indian stock market indices for random walk during April 1996 to June 2001. They found that Indian Stock Market Indices did not follow random walk. Shigguang Ma and Michelle Barnes (2001) tested both Shanghai and Shenzen stock market for efficient market hypothesis using serial correlation, runs and variance ratio test to index and individual share data for daily, weekly and monthly frequencies and found that Chinese stock markets were not weak form efficient. Osei (2002) investigated the asset pricing characteristics and response to annual earnings announcement of the Ghana Stock market. He concluded that Ghana Stock Market is not efficient with respect to annual earnings information releases to the Ghanaian Market. Madhumita Chakraborty (2006) investigated the stock price behaviour using daily closing figures of Milanka Price Index during January 1991 to December 2001 and daily closing prices of twenty-five underlying individuals companies included in the index from July 1991 to May 1999. The study found that stock market in Srilanka did not follow random walk, while results of weak form efficient market hypothesis in twenty-five companies showed mixed outcome. Daniel Simon and Samuel Laryea (2006) examined the weak form of efficient market hypothesis for four African stock markets-Ghana, Mauritius, Egypt and South Africa. Their results implied that South African market was weak form efficient, whereas that of Ghana, Mauritius and Egypt were weak form inefficient.

**OBJECTIVES AND PROPOSED CONTRIBUTIONS:**

Although the literature on the Efficient Market Hypothesis (EMH, Fama (1970)) is very extensive, most studies focus on one type of test of the hypothesis, or one frequency of data, or either individual share or index data. Some researchers do pursue their analysis
using more than one test, or more than one frequency of data, or both individual and index data. However, very few consider all of these elements simultaneously. A careful survey of the existing literature reveals conflicting evidence on weak-form market efficiency for many markets, depending on which test a particular study used, or which type of data the researchers employed. Thus, the question of whether or not Indian Stock Market namely National Stock Exchange is efficient is best answered by a comprehensive and concurrent analysis of the standard tests and various types of data available while using the largest possible sample sizes.

In this paper, we intend to contribute to the empirical literature on tests of Efficient Market Hypothesis by employing various statistical tests to investigate weak form EMH or the random walk hypothesis for National Stock Exchange (NSE) using weekly, monthly and daily frequencies of data set. We also propose to examine the random walk hypothesis in fifty companies of NSE as the study examining the behaviour of the stock prices of individual companies are very rare to the best of our knowledge. A comparison of the results for Indian Stock Market with the results for other countries’ share markets would provide additional understanding of the relative market efficiency of Indian Stock Markets. The New York Stock Exchange in the United States is the largest market in the world, and is generally considered to be a perfectly efficient market in many senses. Thus, tests on the stocks of the New York Stock Exchange are often used as a benchmark to assess the tests on other markets. We will, therefore, use New York Stock Exchange data to compare its results with that of National Stock Exchange’s S&P CNX Nifty data. We also propose to use Chinese Stock Market data set to compare our results with them. The comparison would provide gainful insight in understanding efficiencies in the stock markets.

Our paper deviates from the previous findings of efficient market hypothesis in Indian Stock Market in two ways. First, we would use both parametric and non-parametric techniques to test the validity of EMH in CNX Nifty, its constituents and various indices of NSE. The parametric tests employed here are autoregression, autocorrelation and variance ratio test while non-parametric tests include run test and Kolmogrov-Smirnov normality test. By using more recent data, we would be able to capture the recent trends, which may have impact on the efficiency of financial markets in India.

The findings of this study will be useful to those involved in investment decision-making in the stock market of India, as it will increase their understanding of the pricing process prevailing in the stock market.
A comparison of the results for Indian Stock Market with that of other countries’ share markets especially NYSE and Chinese Stock Market will provide additional understanding of the relative market efficiency of Indian Stock Markets. The comparison of forecasting performances of various models would be useful to judge best fitting model to explain the data in term of forecasting.

**DATA AND METHODOLOGY:**

The data set in our study consists of three sub-samples. One sample would include the daily, weekly and monthly closing prices of S & P CNX Nifty and other indices of NSE for the period January 1991 to April 2007. Sub-sample 2 would comprise of daily closing prices of fifty underlying individual companies included S & P CNX Nifty. The time periods for the second sub-sample vary from stock to stock. Third sample would consist of daily, weekly, monthly closing prices of NYSE and Chinese Stock Market for the period 1991 to April 2007. Data set for first two sub-samples are available on www.nseindia.com, CMIE's prowess and business beacon data base. Data set for third sub sample will be obtainable from www.yahoo.com/finance and www.economagic.com. Lo and Mackinlay (1988) suggest that weekly and monthly data are superior to daily figures since they are free from sampling problems of biases due to bid-ask spreads, non-trading, etc. inherent in the daily prices. Chow and Denning (1993) have stressed that the variance ratio test required a sample size of atleast 256 observations to have reasonable power against other alternative tools. Under these considerations, the weekly observations are the most appropriate data for variance ratio test. In-spite of its limitations, the daily observations would also be included in the study to understand the dynamics.

For the price series, we compute the natural log of the relative price for the given frequency of data set to produce a time series of continuously compounded returns, such that,

\[ R_t = \log(P_t / P_{t-1}) \times 100 \]  

\[ (1) \]

Where \( P_t \) and \( P_{t-1} \) represent the stock price at time \( t \) and \( t-1 \) respectively. A number of complementary procedures for random walks or weak-form market efficiency could be applied. The distribution of the excess returns will be examined for normality. We then examine the autocorrelation function of the data in the markets to ascertain
if one can detect significant relationships among past and present values of the time series properties of the data. If the null hypothesis is rejected on the basis of the performed analysis, we will choose a linear model, which fits the data best in order to analyse its forecasting performance. We will split the data into testing and forecasting sets. We follow the Box-Jenkins approach (1976) for ARIMA building. If no further adjustments to model parameters are required, we then proceed to the forecast analysis. The analysis will be performed on NSE’s S&P CNX Nifty.

The presence of ARCH processes implies that the time series follows some sort of nonlinear dynamic process despite a lack of any significant autocorrelations among past observations. Therefore, if ARCH is present, it may lead to a serious linear model misspecification. For this reason, we test the time series for ARCH effects. Engle (1982) proposed the class of ARCH models, which allowed one to model variance directly in terms of past observations (see Bollerslev, Chou, and Kroner (1992) for an extensive survey of the development and application.

To start with, unit root test of Augmented Dickey Fuller test (ADF) would have been used to determine if the series are difference or trend nonstationary. The study intends to use parametric and non-parametric tests to test the validity of EMH in the market. The brief explanation of each test is given below:

**UNIT ROOT TESTS:**

A unit root test is a statistical test for the proposition that in an autoregressive statistical model of a time series, the autoregressive parameter is one. It is a test for detecting the presence of stationarity in the series. There are three different unit root tests used to test the present of a unit root namely, ADF, Phillips-Perron (PP) test and the Kwiatkowski, Phillips, Schmidt and Shin (KPSS) test. The presence of unit root in a time series is tested with the help of Augmented Dickey-Fuller Test. It tests for a unit root in the univariate representation of time series. For a return series \( R_t \), the ADF test consists of a regression of the first difference of the series against the series lagged \( k \) times as follows:

\[
\Delta r_t = \alpha + \delta r_{t-1} + \sum_{i=1}^{p} \beta_i \Delta r_{t-i} + \epsilon_t
\]  

(2)

\[
\Delta r_t = r_t - r_{t-1}; r_t = \ln(R_t)
\]  

(3)
The null hypothesis is $H_0: \delta = 0$ and $H_1: \delta < 1$. The acceptance of null hypothesis implies nonstationarity. Mackinnon's critical values are used in order to determine the significance of the test statistic. The PP test incorporates an alternative (nonparametric) method of controlling for serial correlation when testing for a unit root by estimating the non-augmented Dickey-Fuller test equation and modifying test statistic so that its asymptotic distribution is unaffected by serial correlation. It is based on the following model:

$$\Delta r_t = \mu + \delta r_{t-1} + \epsilon_t$$  \hspace{1cm} (4)

If the calculated absolute values of ADF and PP statistics are higher than Mackinnon's critical value, we can reject the null hypothesis of random walk in the series.

The KPSS procedure of Kwiatkowski et al (1992) has the advantage of being specifically designed to test the null hypothesis of stationary and unit root as the alternative hypothesis. The test statistics is calculated as:

$$\eta = T^{-2} \sum_{i=1}^{n} S_i^2 / S^2 (L)$$  \hspace{1cm} (5)

Where L is a lag parameter, $S_i$ is the cumulative sum of residuals ($\epsilon_i$) from a regression of the series on a constant a linear trend and where

$$S^2 (L) = T^{-1} \sum_{i=1}^{L} (1 - S)/(L + 1) + \sum_{i=2+1}^{T} \epsilon_i \epsilon_{t-x}$$  \hspace{1cm} (6)

The null hypothesis of stationarity is rejected in favor of the unit root alternative if the calculated statistic exceeds the critical values provided in KPSS (1992).

**NON-PARAMETRIC TESTS:**

**KOLMOGOROV-SMIRNOV (KS) GOODNESS OF FIT TEST:**

The test is used to find out how well a data series fits a particular distribution. Our test compares the cumulative distributional function of the returns with a normal distribution to determine if they are identical. If the probability values for Z for all series turn out to be low, it will confirm that the returns do not conform to normal distribution.

**RUNS TEST:**

The runs test can be used to examine the serial independence in share return movements. The test has an advantage of ignoring the distribution of the data, and does
not require normality or constant variance of the data. The run can be defined as a sequence of return changes of the same sign. An abnormally high number of runs indicate evidence against the null hypothesis of random walk. The significant Z values indicate non-randomness of the series.

PARAMETRIC TESTS:
We intend to employ parametric tests such as autocorrelation test, autoregressive test and variance ratio test to confirm the findings of the non-parametric test and to measure the degree of dependency of the series.

AUTOCORRELATION TEST:
Random walk hypothesis implies independent residuals and a unit root. The independence hypothesis can be investigated by examining the autocorrelation function (ACF). The Autocorrelation tests show whether the serial correlation coefficient are significantly different from zero. In an efficient market, the null hypothesis of zero autocorrelation will prevail. Box-Pierce (1970) gave Q-statistics as an alternative to various hypothesis of autocorrelation with different time lags. It is defined as:

$$LB = T(T+2)\sum_{k=1}^{k} (\hat{\rho}_{k}^2 / T-k)$$

where $\hat{\rho}_{k}^2$ is Autocorrelation Function (ACF) and $T=\text{sample size}$. Q statistics is used to test the validity of market efficiency. It is tested for various values of k. Under the null hypothesis of no autocorrelation, it is distributed as $\lambda^2(k)$. If Q statistics measured found to be significant, it can be said that the market does not follow random walk.

VARIANCE RATIO TESTS:
The robust variance ratio test developed by Lo & Mackinlay (1988) is used on the eight market indices selected for the study for observations from 23-April-96 to 7-June-01. Lo & Mackinlay (1989) have indicated that the variance ratio test is more powerful than the well-known Dickey-Fuller unit root or the Box-Pierce Q tests. The first null hypothesis is stated as follows:

**H0:** The variance ratio at lag q is defined as the ratio of the variance of the q-period return to the variance of the one-period return divided by q, which is unity under the random walk hypothesis:

$$\text{i.e. } VR(q) = \frac{\text{Var}[r(q)]}{q \cdot \text{Var}[r]} = 1 \quad (7)$$
The alternative hypothesis will be \( VR(q) \) is not equal to one.

\[
VR(q) = \sigma^2(q) / \sigma^2(1)
\]

where \( \sigma^2(q) \) is \( 1/q \) the variance of the \( q \)-differences and \( \sigma^2(1) \) is the variance of the first difference. Where,

\[
\sigma^2(q) = 1/m \sum_{i=q}^{mq} (Y_i - Y_{i-q} - \mu)^2 \quad \text{and} \quad \sigma^2(1) = 1/nq - 1 \sum_{i=1}^{mq} (Y_i - Y_{i-1} - \mu)^2
\]

where

\[
\mu = 1/nq(Y_{mq} - Y_0) \quad \text{and} \quad Y_o \quad \text{and} \quad Y_{mq} \quad \text{are the first and last observations of the logarithmic price series.}
\]

\[m = q(nq-q+1)(1-1/n)\]

After deriving an asymptotic distribution of the variance ratios, two alternative test statistics are derived to test the null hypothesis for different specifications of error term behavior. The first test statistic, \( Z(q) \), assumes an independent and identical distributed normal error term. Then, the standard normal test statistic is computed as follows:

\[
Z(q) = [VR(q) - 1]/[\varphi(q)^{1/2}] \sim N(0,1)
\]

Where, \( \varphi(q) = (2(2q-1) (q-1)) / [3q(nq)] \)

The second test statistic, \( Z^*(q) \), allows for a general heteroscedasticity of error term. The heteroscedasticity consistent standard normal test statistics relaxed the assumption of normality. The formula is given as follows:

\[
Z^*(q) = [VR(q) - 1]/[\varphi^*(q)^{1/2}] \sim N(0,1)
\]

Then, \( \varphi^*(q) \), the heteroscedasticity consistent asymptotic variance of the variance ratio is computed as follows:

\[
\varphi^*(q) = \theta / nq
\]

where, \( \theta = 4 \sum_{k=1}^{q-1} [1 - 1 \{1 - (k/q)\} \delta_k] \)

\[
\delta_k = \left[ nq \sum_{j=k+1}^{mq} (Y_j - Y_{j-1} - \mu)^2 (Y_{j-k} - Y_{j-k+1} - \mu)^2 \right] / \sum_{j=1}^{mq} (Y_j - Y_{j-1} - \mu)^2
\]

Both statistics \( Z(q) \) and \( Z^*(q) \) are shown to be asymptotically standard normal.

**Multiple Variance Ratio Test:**
Chow and Denning (1993) proposed the test to detect autocorrelation and heteroscedasticity in the returns. It generates the procedure for multiple comparison of the set of variance ratio estimates with unity. For a single variance ratio test, under null hypothesis, \( \text{VR}(q)=1 \), hence \( \text{Mr}(q)=\text{VR}(q)-1=0 \). Under the random walk null hypothesis, there are multiple sub-hypothesis:

\[ H_0 = \text{Mr}(q)=0 \text{ for } i=1,2,\ldots,m. \]
\[ H_{1i}: \text{Mr}(q) \neq 0 \text{ for any } i=1,2,\ldots,m. \]

The rejection of any one or more \( H_0i \) rejects the random walk null hypothesis.

**Autoregression Test:**
Using maximum likelihood auto-regression techniques, we attempt to confirm the results of the variance ratio test by testing if there is a non-zero significant relationship between the current return series and its first and second lags. The coefficients will be examined to find out their statistical significance. If the coefficients are statistically not different from zero then market can be said to be an efficient market.

**Forecasting Models:**
If the null hypothesis is rejected on the basis of the performed analysis, we will choose a linear model, which fits the data best in order to analyse its forecasting performance. We will split the data into testing and forecasting sets. We follow the Box-Jenkins approach (1976) for ARIMA building. If no further adjustments to model parameters are required, we will then proceed to the forecast analysis.

The presence of ARCH processes implies that the time series follows some sort of nonlinear dynamic process despite a lack of any significant autocorrelations among past observations. Therefore, if ARCH is present, it may lead to a serious linear model misspecification. For this reason, we will test the time series for ARCH effects. Engle (1982) proposed the class of ARCH models which allowed one to model variance directly in terms of past observations (see Bollerslev, Chou, and Kroner (1992) for an extensive survey of the development and application of these models). These models represent conditional variance as a distributed lag of past squared innovations:

\[ \varepsilon \sim N(0,h_t) \]

\[ h_t = \alpha_0 + \sum_{j=1}^{q} \alpha_j \varepsilon_{t-j}^2 \quad (8) \]
where $\varepsilon_t$ represents past innovations, $q$ denotes the lag operator, and $\alpha > 0$, $q \geq 0$.

To avoid estimation of a large number of coefficients in a high-order polynomial, Bollerslev (1986) generalizes the order of the ARCH($q$) model into the Generalised Autoregressive Conditional Heteroskedasticity model of orders $p$ and $q$:

$$\varepsilon \sim N(0, h_t)$$

$$h_t = \alpha_0 + \sum_{j=1}^{q} \alpha_j \varepsilon_{t-j}^2 + \sum_{i=1}^{p} \beta_i h_{t-i}$$ (9)

This process is covariance stationary if and only if $(\alpha_1 + \alpha_2 + \ldots + \alpha_q) + (\beta_1 + \beta_2 + \ldots + \beta_p) < 1$, in which case the model could also be written as an infinite order ARCH model.

We will choose the best fitting AR($p$) model and perform an extensive analysis of ARIMA model residuals and squared residuals. The following tests will be conducted: Lagrange multiplier test for residual serial correlation; Ramsey’s RESET test for functional form; Jarque-Bera normality test based on a test of skewness and kurtosis of residuals; and a residuals test for the presence of ARCH effects based on a regression of squared residuals on squared fitted values.

Moreover, we will analyse the autocorrelation function of the linear model residuals. If there is no significant autocorrelation, we will proceed to the analysis of ACF and PACF of the squared residuals. As McLeod and Li (1983) show, the Box-Pierce Q-statistics of the squared residuals of an ARMA model can be used to test for GARCH identification. We will choose the $p$, $q$ orders of the GARCH process using AIC and SBC criteria for the set of all possible combinations of $p$, $q$. The model with the highest values of AIC and SBC will be chosen for further analysis. Then we will select the type of the GARCH process, which would best fit the analysed series. We will then proceed to build ARCH, GARCH, GARCH-M, TGARCH, and EGARCH models and estimate their AIC and SBC criteria. The GARCH-M model explains risk-return relationship in the market while TGARCH and EGARCH will be used to capture leverage effect or asymmetric effect in the stock market. The model with the maximum value for these criteria fits the data best and will be chosen for further analysis. Finally, the forecast will be performed and results will be analysed. A dynamic forecasting approach is used, in which the testing set will be enlarged successively by the forecast values. The accuracy of forecasts could be estimated using Mean Prediction Error, Sum of Squares of Prediction Error, Root Mean Sum of Squares of Prediction Error and Mean Sum of Absolute Prediction Error. This will help us to find out comparatively accurate forecasting models in S&P CNX Nifty that can be used for forecasting and for trading to devise profitable strategy or to take investment decision.
SOFTWARE PACKAGE TO BE USED:

I will use Eviews5 and SPSS 15 for the data analysis of the proposed study.

REFERENCES:


