IOSR Journal of Economics and Finance (IOSR-JEF)
e-ISSN: 2321-5933, p-ISSN: 2321-5925.Volume 14, Issue 1 Ser. III (Jan. – Feb. 2023), PP 37-43
www.iosrjournals.org

Stylized facts of Nifty return series

Dr. Sonal Sharma[1]
[1]Hansraj College, University of Delhi

Abstract
Stylized facts indicate the traditionally found expected patterns in the time series of an asset. Present study tries to investigate the stylized facts in the daily Nifty Index returns. We tested the Nifty time series for fat tails, symmetry, serial correlation and volatility clustering by taking data from 2002 to 2021.It was found that the Nifty series has negative skewness with longer tails than those of the normal distribution. The correlogram of the log returns indicated no linear autocorrelation. The autocorrelation functions of NIFTY absolute and squared returns indicated towards long range dependence in volatility. The preliminary identification tests performed on the log-returns data indicated normality of the return distribution. Lastly, the two unit root tests, viz., the Augmented Dicky-Fuller (ADF) test and the Phillips-Perron (PP) tests indicated that the NIFTY log return series is stationary. Moreover, the Ljung-Box Q-Statistic indicated that the NIFTY log return series is not independently distributed.

Date of Submission: 03-02-2023 Date of Acceptance: 15-02-2023

I. Introduction

Bachelier in 1900 proposed for the first time that the price changes between transactions are independent random variables. They can be assumed to be following stochastic process and therefore the random walk theory was modelled later on by Fama (1965). Literature shows there are innumerous models available to model stock returns. "All these stock return models have a common target, that is to approximate the behavior of the unobservable data generating process (DGP) that determines observed stock prices. Assessing the adequacy of this approximation is where the "stylized facts" of stock price movements come in". (Thompson, 2011).

Any variable when observed over a period of time at constant intervals is known as time series. When a time series of a variable say stock prices is plotted over a period of time, it may show an intriguing pattern. One could reason out these patterns by introducing outside explanatory variables. Sometimes the variable's history itself explains the patterns. The latter approach leads to time series modeling. It is therefore very crucial to identify these patterns from a sample of the underlying asset's prices over a period of time. The aim of the present study is to study the relevant literature and identify the stylized facts of the Nifty return series in order to make a comparison with the same.

Stock returns are generally assumed to follow a logarithmic diffusion process in continuous time with constant drift and volatility parameters, as shown in equation (1).

$\frac{dS}{S} = \mu dt + \sigma dz$……………….[1]

Starting from an initial value S_0, the log return over the non-infinitesimal period from 0 to T is given by

$R = \ln(S_T / S_0)$ ………………..[2]

Definite patterns in the time series graph of these returns over a period of time can be identified because actual security prices don't follow equation (1).Different information sets affect the time series of various assets differently and therefore these series exhibit different properties.But when examined from statistical point of view, the apparently random variations of asset prices do share some quite non-trivial statistical properties. Such particularities are commonly established asstylized empirical facts. They are common through a wide range of instruments, markets and time periods.

The classical theory of finance relies on the assumption that asset returns are identically and independently distributed (iid). The normality assumption when added leads to the Brownian motion behavior (Osborne 1959).

2 Patterns in Financial time series

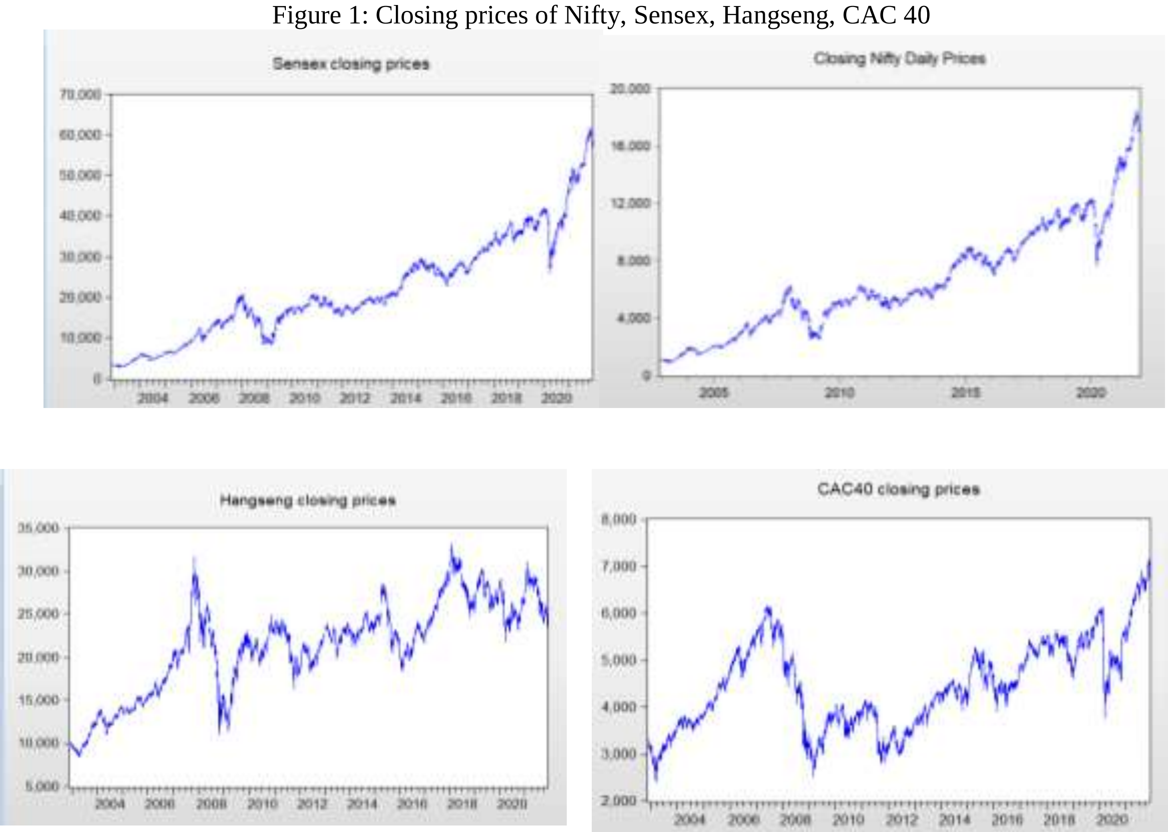
Financial modeling is all about seizing and manipulating patterns in the financial time series data. In this section the unique statistical properties of financial market data from the Nifty Index and several so-called stylized facts are discussed. The nifty closing prices are taken for a period starting from December 1, 2002 to December 1, 2021.

Do various asset prices like commodity prices, stock prices or exchange rate prices behave in a particular fashion? They are not expected to, since they are affected by different and varied reasons and information sets. However, many empirical studies, for example, Bollerslev, Chou, and Kroner (1992), Pagan (1996), Campbell, Lo and Mackinlay (1997), Cont (2001), and Gourieroux and Jasiak (2001) have identified some common attributes among financial data that are acknowledged as stylized facts. Cont (2001), in particular, provides a comprehensive survey of these stylized facts, some of which include and are tested for the Nifty return series as follows:

(A) Variation across different assets:

To give a brief appreciation of the amount of variation possible across various financial Indices figure 1 plots the closing prices of the Nifty index and for a comparison also shows the closing prices of some other indices like Sensex, Hongkong's Hangseng Index, and France's CAC 40 Index.

Figure 1: Closing prices of Nifty, Sensex, Hangseng, CAC 40



There has been relatively more variation in Sensex than Nifty. There has been more pronounced fall in Hangseng and CAC 40 as compared to CAC 40and Nifty around the crises of 2008. Over the whole period it can be seen that Hangeng and CAC 40 has more variations as compared to Sensex and Nifty. Thus, various asset prices show different variations over time because different set of factors operate in these markets, thereby creating different levels of variations in the asset prices.

(B) Structure of the time series:

Structure of a time series can be analyzed through the concept of moment. Mean is the first moment which is also taken to be a measure of location. This stands true if the return distribution is a symmetric one. If the distribution is not symmetric then median, which is the 50^{th} percentile of the sample, is taken to be more robust against outliers. So if one large value is part of the sample then it will move the mean away from the central part but won't impact the median. The mean and the median being an integral part of the descriptive statistics reported for any research is mentioned for the Nifty sample return series in figure 2.

Variance or the standard deviation is a popular measure of dispersion of the return series. Volatility of the return series is often measured through the standard deviation in finance. Standard deviation has the same unit of measure as the mean and thus is more convenient to use than variance. Standard deviation of the Nifty return series is given in figure2.

The symmetry of a distribution can be analyzed through its skewness. If large negative realizations of the variable are more likely to appear, then skewness of such variable's distribution will be negative. Positive skewness will result if large positive realizations are more likely to appear then negative ones. A zero skewness is an indication of a normal distribution. Skewness of the Nifty return series, given in figure 2 is -0.4946 indicating a negative skewness with large negative realizations to occur more likely.

The most popular measure to know the structure of a time series is the tail thickness, and thereforenon-normality, of the distribution which can be measured through kurtosis.Kurtosis of the sample returns can be defined through following equation:

$$k = \frac{1}{n-1}\sum_{i=1}^{n}\frac{(R_i-\bar{R})^4}{S^4} \qquad \ldots\ldots\ldots\ldots(3)$$

Where $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(R_i-\bar{R})^2$ and $\bar{R} = \frac{1}{n}\sum_{i=1}^{n}R_i$ and n is the number of observations in the return set taken as sample.Most of the unconditional return distributions of financial variables have fat tails than a normal distribution. Normal distribution has kurtosis of 3.

Fama in 1965 identified leptokurtosis in each of the thirty daily stock returns included in Dow Jones Industrial Average. Ding, Granger and Engle (1993) test the daily return series of S&P 500 and estimated the skewness as -0.487 and kurtosis as high as 25.42. More recent studies also show similar results. For example, Bergsli, etal (2022) studies the daily log return series of Bitcoin data from 2011 to 2018. The authors report the skewness as -0.335 and kurtosis of the series as 8.8. The literature above shows that it is now well accepted that most of the series of stock returns exhibit leptokurtosis and non-normality. Present study analyzes the Nifty return series from 2002 to 2021 and the results are shown in figure 2. It can be seen that the daily Nifty return series has a kurtosis of 15.407 which is more than 3 indicating that the series is leptokurtic.

Figure 2: Histogram of Daily Log return series of Nifty index

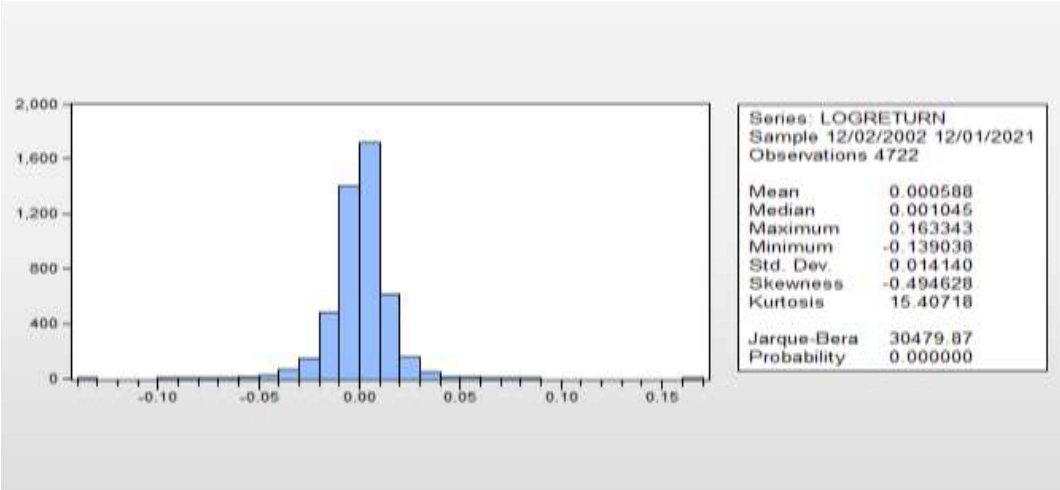
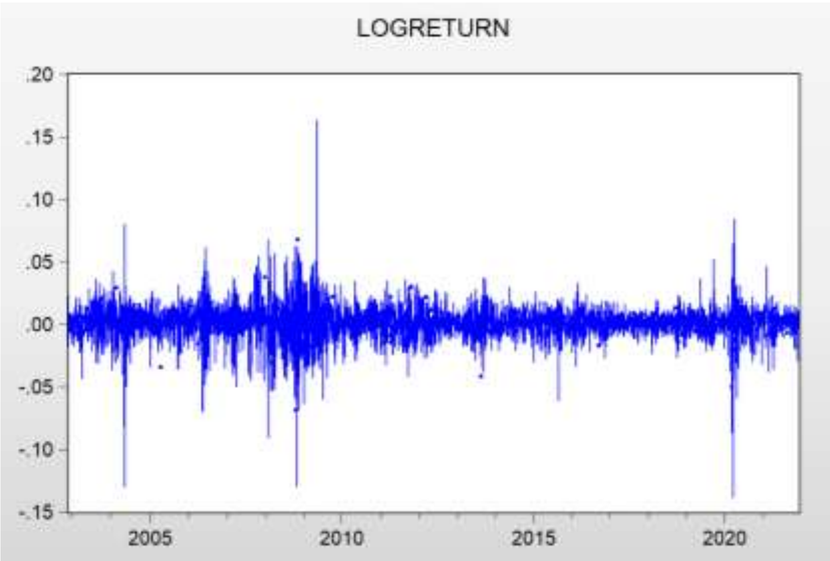
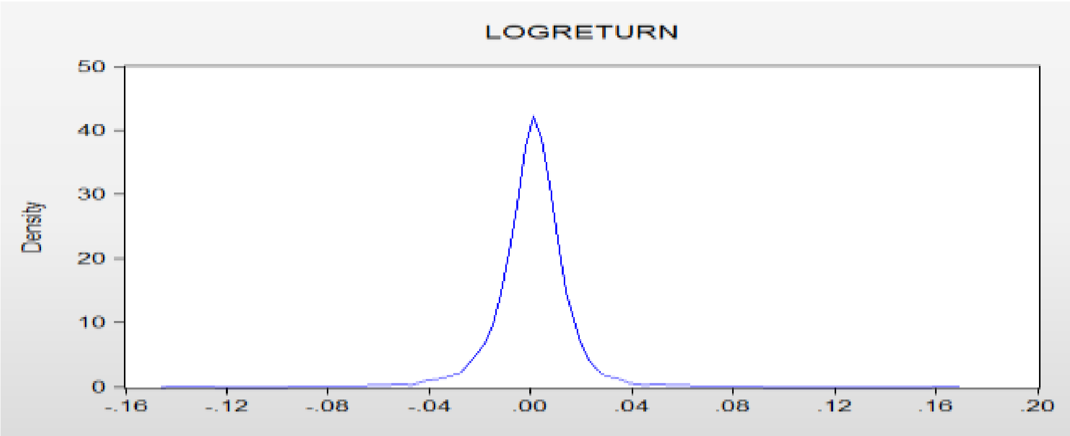


Figure 3: Nifty log return series



Histogram of the return series divides the total distance between the minimum and the maximum values into a number of equal length intervals and shows the number of observations lying in each interval. The histogram of the Nifty return series is displayed in figure 2. As we can see most of the returns lie near mean which is almost zero (0.000588). There are more of positive returns as compared to negative ones.The log return series of the Nifty returns is depiceted in figure 3.In addition to histograms , sometimes the kernel density functions are also reported by researchers. The kernel density function graph is shown in figure 4.

Figure 4: Kernel Density function of Log return series of Nifty Index



(C) Normality of return distribution

Normality can be tested through the Jarque-Bera test which is based on moments of the return series. The Jarque-Bera test is a test to measure the departure of the return distribution from normality, based on the sample kurtosis and skewness. The test statistic measures the difference of the skewness and kurtosis of the series with those of the normal distribution. The JB test statistics can be defined as

$$JB = \frac{n}{6}[S^2 + \frac{(K-3)^2}{4}] \qquad \text{...............[4]}$$

Where n is the number of observations (or degrees of freedom in general); S is the sample skewness, K is the sample kurtosis. The JB statistic has an asymptotic chi-square distribution andtries to examine the null hypothesis of data belonging to a normal distribution. The null hypothesis is a joint hypothesis of zero skewness and no excess kurtosis. The stated probability is the probability that a JB statistic surpasses (in absolute value) the observed value. Atrivial probability value clues to the rejection of the null hypothesis of a normal distribution. From figure 2 it can be seen that, the JB statistics is very high for the Nifty return series. The p-values are also zero, thereby implying non-normality of the return distribution, both at 5% as well as 1% significance level. Thus, it can be concluded that Nifty's daily returns, during 2002-2021, are not normally distributed but are leptokurtic (as kurtosis exceeds 3) and negatively skewed.

(D) Linear Dependence in returns: Autocorrelation of Returns

Autocorrelation means correlation of a variable with a deferred observation of itself. An extensively popular characteristic of daily return series in literature is non-appearance of significant linear autocorrelation. The autocorrelation in a sample series of daily returns can be calculated as:

$$\rho_{t,R} = \frac{\sum_{i=1}^{n-t}(R_i-\overline{R})(R_{i-t}-\overline{R})}{\sum_i^n (R_i-\overline{R})^2} \qquad \text{..........................(5)}$$

"It is now well established that the stock market returns themselves contain little serial correlation [Fama (1970), Taylor (1986)] which is in agreement with the efficient market theory. But this empirical fact does not necessarily imply returns are independently identically distributed as many theoretical financial models assume. It is probable that the series is though serially uncorrelated but is dependent. The stock market data is especially so since if the market is efficient, a stock's price should change with the arrival of information. If information comes in bunches, the distribution of the next return will depend on previous returns although they may not be correlated" Ding, etal (1993). Bartlett (1946) has demonstrated that if log return data is representing an independent and identical process then the sample autocorrelation of the return series is approximately N(0, l/T). Ding, etal (1993) displayed the first lag autocorrelation to be significantly positive. Like Ding, etal many researchers have established that most stock market return series have a very small positive first order autocorrelationthereby suggesting that the return series do have some memory although it is very short. This guides to the conclusion that the efficient market or random walk hypothesis does not hold strictly. Significant first order autocorrelation has originated in illiquid or less developed markets. For example, Yu (2002) states "While higher order autocorrelations are in general diminishing, the first autocorrelation is low but not negligible. This is the evidence of volatility clustering and suggests that the volatility is predictable." Present study deals with Nifty Index return series from India and the autocorrelations of log return series at various lags are reported in table 1 (a). As can be seen from the table that the autocorrelations at various lags linger around zero, thereby demonstrating that it is a stationary time series.

(E) Non-linear dependence in returns

Absence of autocorrelations in daily return series loaded some empirical support for "random walk" model of prices with independent random returns. However, it is brimming that the absence of serial correlation does not pinpoint the independence of the increments.Independence connote that any nonlinear function of returns will also have no autocorrelation which generally does not hold. Thus, the simple nonlinear functions of returns, such as absolute returns or squared returns, exhibit compelling positive autocorrelation. This is prominently known as volatility clustering, that is, large price disparities are more likely to be tailed by large price disparities. Log price returns are consequently not random walks. Volatility clustering can be studied by calculating the autocorrelation function of the squared returns defined as:

$C_2(\tau) = corr(|r(t+\tau,\Delta t)|^2, |r(t,\Delta t)|^2)$ …………[6]

Literature shows that this autocorrelation function hover positive and decays slowly, remaining significantly positive over several days, sometimes weeks. This is frequently called the "ARCH effect" in the econometric literature. This persistence can be used as it has some degrees of predictability for the amplitude of the returns as evaluated by their squares. In the same way, the autocorrelation functions of distinct powers of returns can be considered:

$C_\alpha(\tau) = corr(|r(t+\tau,\Delta t)|^\alpha, |r(t,\Delta t)|^\alpha)$. …………[7]

The autocorrelation function of the absolute returns can be defined as:

$C_0(\tau) = corr(|r(t+\tau,\Delta t)|, |r(t,\Delta t)|)$. …………[8]

The autocorrelation functions for both the absolute as well as squared returns of Nifty are shown below in table 1 (b) and (c) respectively.

Table 1: a)ACFs and PACFs of Log returns

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.048	0.048	10.731	0.001
		2	-0.028	-0.031	14.521	0.001
		3	-0.001	0.002	14.524	0.002
		4	-0.000	-0.001	14.525	0.006
		5	0.007	0.007	14.747	0.012
		6	-0.065	-0.066	34.863	0.000
		7	0.036	0.043	41.037	0.000
		8	0.031	0.024	45.680	0.000
		9	0.008	0.008	45.988	0.000
		10	0.037	0.038	52.465	0.000
		11	-0.031	-0.033	57.001	0.000
		12	0.016	0.016	58.094	0.000
		13	0.019	0.020	59.750	0.000
		14	0.035	0.036	65.453	0.000
		15	0.006	0.001	65.601	0.000
		16	-0.014	-0.008	66.557	0.000
		17	0.046	0.040	76.641	0.000
		18	-0.022	-0.026	78.994	0.000
		19	-0.002	0.006	79.010	0.000
		20	-0.031	-0.032	83.541	0.000

b) ACFs and PACFs of Absolute returns

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.284	0.284	382.13	0.000
		2	0.301	0.240	811.47	0.000
		3	0.281	0.171	1185.1	0.000
		4	0.293	0.159	1590.1	0.000
		5	0.264	0.103	1919.0	0.000
		6	0.272	0.103	2267.7	0.000
		7	0.258	0.079	2583.5	0.000
		8	0.238	0.047	2850.4	0.000
		9	0.261	0.080	3173.2	0.000
		10	0.256	0.068	3482.8	0.000
		11	0.207	0.001	3685.8	0.000
		12	0.201	0.004	3877.9	0.000
		13	0.203	0.015	4072.6	0.000
		14	0.218	0.042	4298.2	0.000
		15	0.215	0.040	4517.3	0.000
		16	0.199	0.017	4704.6	0.000
		17	0.182	0.001	4861.2	0.000
		18	0.194	0.023	5039.3	0.000
		19	0.179	0.005	5190.8	0.000
		20	0.193	0.030	5367.4	0.000

c) ACFs and PACFs of Squared returns

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.215	0.215	217.47	0.000
		2	0.217	0.179	440.16	0.000
		3	0.147	0.076	541.84	0.000
		4	0.170	0.102	678.67	0.000
		5	0.178	0.105	828.98	0.000
		6	0.127	0.033	905.52	0.000
		7	0.189	0.109	1074.2	0.000
		8	0.095	-0.006	1117.3	0.000
		9	0.192	0.109	1291.6	0.000
		10	0.158	0.064	1409.4	0.000
		11	0.113	-0.004	1470.1	0.000
		12	0.089	-0.012	1507.5	0.000
		13	0.094	0.014	1549.4	0.000
		14	0.105	0.013	1601.7	0.000
		15	0.109	0.036	1658.4	0.000
		16	0.099	0.008	1704.9	0.000
		17	0.072	-0.009	1729.3	0.000
		18	0.092	0.022	1769.3	0.000
		19	0.060	-0.020	1786.3	0.000
		20	0.075	0.007	1812.8	0.000

(F) Stationarity

Statistical analysis of any data becomes advantageous only ifpartially some statistical properties of the data remain stable overtime. Stationary process is a stochastic process whose joint probability distribution doesn't alter when changed in time or space. A stochastic process will be labelled as stationary if its mean and variance are constant over time. Moreover, the covariance between the two time periods hinge only on the lag between the two time periods and not the actual time at which the covariance is computed. In literature, such a stochastic process is popularly christened as weakly or second-order stationary, covariance stationary or simply a stochastic process. Time series data generally is transformed to become stationary through differencing, first-order or may be higher-order. One requires stationarity of the data so that inferences can be made from the sample set.

There are two general methodologies to test for stationarity, namely, parametric and nonparametric. Parametric tests involve making certain assumptions as respects the nature of the data. Nonparametric tests are not based on such assumptions and therefore are more widely endorsed. The nonparametric tests include the Runs test, Phillips-Perron test, etc.

The stationarity of the NIFTY return series was tested through the the Augmented Dickey-Fuller (ADF) test and the Phillips and Perron (PP) test. The two statistics scrutinizes whether the process has a unit root or not, i.e. a problem of non-stationarity. For a return series R_t the ADF test embody a regression of the first difference of the series against the series lagged k times as follows:

$$\Delta r_t = \alpha + \delta r_{t-1} + \sum \beta_s \Delta r_{t-s} + \epsilon_t \qquad \text{……………[9]}$$

Where $\Delta r_t = r_t - r_{t-1}$; $r_t = \ln(R_t)$. The null and alternate hypotheses are $H_0: \delta = 0; and\ H_1: \delta < 1$. The affirmation of the null hypothesis implies nonstationarity. The PP statistic is an alternative test of $\delta = 0$. There is no lagged difference term and it is based on the ensuing model:

$$\Delta r_t = \mu + \delta r_{t-1} + \varepsilon_t \qquad \text{……………[10]}$$

Table 2: (a) ADF test for Nifty log-return series

Null Hypothesis: LOGRETURN has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic - based on SIC, maxlag=31)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-65.49742	0.0001
Test critical values:	1% level	-3.431553	
	5% level	-2.861956	
	10% level	-2.567034	

*MacKinnon (1996) one-sided p-values.

(b) PP test for Nifty log-return series

Null Hypothesis: LOGRETURN has a unit root
Exogenous: Constant
Bandwidth: 7 (Newey-West automatic) using Bartlett kernel

		Adj. t-Stat	Prob.*
Phillips-Perron test statistic		-65.43172	0.0001
Test critical values:	1% level	-3.431553	
	5% level	-2.861956	
	10% level	-2.567034	

*MacKinnon (1996) one-sided p-values.

Residual variance (no correction)	0.000199
HAC corrected variance (Bartlett kernel)	0.000187

The results for both the ADF and the PP test statistics are presented in table 2 (a) and (b) for the Nifty log return series respectively. T-statistic (or tau (τ) statistic) is computed for both the tests and then compared with the MacKinnon critical values. If the calculated value is more than the MacKinnon critical value, than we affirm the hypothesis that $\delta = 0$, in which case the time series is non-stationary. And if the calculated value is less than the MacKinnon critical value, than we decline the hypothesis that $\delta = 0$, in which case the time series is stationary. The ADF statistic value for NIFTY in part A of table 2 is -65.49742 and the critical values at the 1%, 5% and 10% levels are all higher than the ADF calculated statistic and therefore the null hypothesis that the NIFTY log return series has a unit root is declined. In other words, the series is stationary. Similarly, the Phillips-Perron (PP) test results shows a statistic of -65.43172, which is less than the critical values at all significance levels, thereby affirming the series to be stationary. Thus, it can beassured through both the test, i.e., the ADF and the PP test statistics, that the Nifty log return series is a stationary series over the period of study.

II. Conclusion

The present study is a short description of various stylized facts about the NIFTY index return distribution. A detailed analysis of the structure of the time series of log returns was done. The analysis, in the form of descriptive statistics, leads to a number of conclusions. Firstly, it can be seen that the average (mean) of the NIFTY returns is positive, thereby indicating that the price series have increased over the period of study, i.e., from December 2002 to December 2021. Secondly, a negative skewness implies asymmetry with a heavier left tail. A large positive kurtosis indicates towards longer tails than those of the normal distribution. The analysis of the dependence properties of the Nifty returns indicates that there is no linear autocorrelation, as is judged through the correlogram of the log returns. The non-linear dependence in the NIFTY returns can be judged through the autocorrelation function of the absolute and squared returns. Both the autocorrelation functions decay slowly, remaining significantly positive over several days which implies some degree of predictability and a sign of long range dependence in volatility. After identifying the stylized facts of the NIFTY returns, some preliminary identification tests were performed on the log-returns data. Firstly, a normality test called the Jarque-Bera test was conducted. The test result indicates a rejection of the null hypothesis of normality of the return distribution both at 1% and 5% level of significance. Thus, it can be seen that the daily NIFTY returns were not normally distributed but are leptokurtic and negatively skewed. Secondly, stationarity of the returns was tested to understand whether inferences from the sample set can be made or not. This was done, firstly, through examination of the autocorrelation function, i.e., the correlogram diagram and then, through the two unit root tests, viz., the Augmented Dicky-Fuller (ADF) test and the Phillips-Perron (PP) test. Both the ADF test and the PP test, as well as the correlogram indicated the same fact, that is, the NIFTY log return series is a stationary. Third and lastly, the time dependency of returns was judged through the Ljung-Box Q-Statistic test. The Q-Stat indicates that the NIFTY log return series is not independently distributed.

References

[1]. Bachelier, L. (1900). “Theory of speculation”. In Annales scientifiques de l'École normale supérieure (Vol. 17, pp. 21-86).
[2]. Bartlett, M. S. (1946). “On the theoretical specification and sampling properties of autocorrelated time-series”. Supplement to the Journal of the Royal Statistical Society, 8(1), 27-41.
[3]. Bergsli, L. Ø., Lind, A. F., Molnár, P., & Polasik, M. (2022). “Forecasting volatility of Bitcoin”. Research in International Business and Finance, 59, 101540.
[4]. Bollerslev, T., Chou, R. Y., & Kroner, K. F. (1992). “ARCH modeling in finance: A review of the theory and empirical evidence”. Journal of econometrics, 52(1-2), 5-59.
[5]. Campbell, J. Y., A. Lo, and A. C. MacKinlay, (1997). The Econometrics of Financial Markets (Princeton University Press, Princeton, New Jersey).
[6]. Cont, R. (2001). “Empirical Properties of Asset Returns: Stylized Facts and Statistical Issues” Quantitative Finance, 1, 223-236.
[7]. Ding, Z, Granger, C., and Engle, R. F. (1993). “A long memory property of stock market returns and a new model”. Journal of Empirical Finance, 1, 83-106.
[8]. Fama, E. (1965). “The behaviour of stock market prices”. Journal of Business, 38, 34-105.
[9]. Fama, E. F. (1970). “Efficient capital markets: A review of theory and empirical work”. The journal of Finance, 25(2), 383-417.
[10]. Gourieroux, C., & Jasiak, J. (2011). Nonlinear persistence and copersistence (pp. 77-103). Palgrave Macmillan UK.

[11]. Osborne. M.F.M. (1959) "Brownian Motion in the Stock Market". Operations Research, 7(2), 145- 173
[12]. Pagan, A., (1996). "The econometrics of financial markets". Journal of Empirical Finance, 3, 15-102.
[13]. Taylor, S. J. (1986): Modeling Financial Time Series. Chichester: John Wiley & Sons.
[14]. Thompson, S. (2011). "The stylised facts of stock price movements". The New Zealand Review of Economics and Finance, 1.
[15]. Yu, J. (2002). "Forecasting volatility in the New Zealand stock market". Applied Financial Economics, 12(3), 193- 202.

XXXXXXXXX. "Stylized facts of Nifty return series." *IOSR Journal of Economics and Finance (IOSR-JEF)*, 14(1), 2023, pp. 37-43.