The International Journal of Applied Economics & Finance

ISSN 1991-0886
Study of Stylized Facts in Indian Financial Markets

Indranil Mukherjee, Chitrakalpa Sen and Amitava Sarkar
School of Management and Sciences, West Bengal University of Technology, BF 142, Sector I, Salt Lake, Kolkata-700064, India

Corresponding Author: Indranil Mukherjee, School of Management and Sciences, West Bengal University of Technology, BF-142, Sector-I, Salt Lake, Kolkata-700064, India Tel: 0913323210731 Ext 203 Fax: 0913323341028

ABSTRACT

Stylized empirical facts emerging from the statistical analysis of price variations in various types of financial markets have attracted the attention of researchers since a long time. These are a set of properties, common across many instruments, markets and time periods that has been observed by independent studies. The objective of this research is to study stylized facts of financial time series by using data from the Indian financial market. BSE SENSEX is used as a proxy for the Indian market. Particular stress is given to the study of volatility using different models from the GARCH family including GARCH, EGARCH, TARCH, Asymmetric Component GARCH etc. The raw SENSEX daily return series are found to be non-normal having fat tails, show significant amount of asymmetry and exhibit strong volatility persistence as well as volatility clustering. This study also examines the possibility of long-term dependence (long memory) in Absolute SENSEX daily return series. Rescaled range analysis, modified rescaled range analysis and GPH test are used for this purpose. The results indicate presence of long memory and also show the series to be fractionally integrated. The findings obtained are in broad agreement with the stylized facts observed in financial time series.

Key words: Stylized facts, volatility, GARCH models, long memory, R/S analysis, fractional integration

INTRODUCTION

Financial time series exhibit some apparently puzzling empirical (statistical) regularities. These are known as stylized facts. Stylized facts are empirical observations that are so consistent and have been made in so many contexts (e.g., across a wide range of instruments, markets and time periods) that they are accepted as truths, to which theories must fit (Cont, 2001, 2008). Thus stylized facts are universal regularities, independent of time, place and specific compositional details. Stylized facts are obtained by taking a common denominator among the properties observed in studies of different markets and instruments. Due to their generality, stylized facts are often qualitative and not quite precise enough to distinguish among different parametric models (Coonen, 2004; Ding et al., 1993).

The universality of the regularities indicates that the stylized facts can be thought to have some common origin. Some of the commonly occurring stylized facts are given below (Brock and de Lima, 1995; Campbell et al., 1996):

• Absence of autocorrelation-(linear) autocorrelations of asset returns are not at all significant; analyses of high frequency data show that correlation times can be as short as a few minutes in highly traded stocks or indices
Slow decay of autocorrelation in absolute returns—the autocorrelation function of absolute returns decays slowly as a function of the time lag, approximately as a power law; this fact is often interpreted as a sign of long-range dependence.

Fat tails—the (unconditional) distribution of returns appears to display a power-law or pareto-like tail, with a tail index which is finite and greater than two (positive excess kurtosis). The distributions are approximately bell-shaped but assign more than normal probability to events in the center (more peaked) and at the extremes (exhibit heavy tails).

Excess volatility—several empirical studies indicate that it is not easy to justify the observed level of variability in asset returns by taking into account variations in the fundamental economic variables. The fact that large (positive or negative) returns occur cannot always be explained by the arrival of new information on the market.

Volatility clustering—different measures of volatility display a positive autocorrelation over several days, which quantifies the fact that high and low volatility events tend to cluster in time; in other words, periods of intense fluctuations and mild fluctuations tend to cluster together.

Volatility persistence (long memory)—there is dependence between stock market returns at different times; technically, the volatility has slowly decaying autocorrelations and there is nonlinear dependence.

The stylized facts show that the normal distribution is inadequate in describing the distribution of the financial time series. It is necessary to invoke a distribution which assigns more probability to the tails of the distribution and thus, immediately alludes to power-law type distributions (a strong signature of complexity). The appearance of these facts in apparently unrelated and diverse systems suggests the existence of some common underlying mechanism responsible for the appearance of these facts (collective behaviour and critical phenomena typical of complex systems). Thus the existence of stylized facts in financial markets reinforces the idea of the latter as complex systems (Borland, 2005; Johnson et al., 2003). Furthermore, the dependence properties of asset returns and the phenomenon of volatility clustering have led to the development of stochastic models in finance—the GARCH School of models is used essentially to model these phenomena.

In this communication, we address the issue of stylized facts by using data from the Indian stock market. While descriptive statistics are studied briefly, a great deal of emphasis is put on the study of volatility by applying several models from the GARCH school like GARCH, EGARCH, TARCH and Asymmetric Component GARCH model. We also examine the probable existence of long memory in financial time series. This is an empirical investigation using statistical tools to show the relevance of the stylized facts and other related issues.

This study is unique for the following reasons:

- Several authors have carried out studies on stylized facts, but there has been little study using data from the Indian financial market.
- This is probably the first systematic and exhaustive study to verify the existence of stylized facts for the Indian financial market. The Indian financial market is an emerging one (the Indian economy opening up essentially from the mid-nineties upwards) and this study captures the entire spectrum from 1997 to 2009 (when the study was concluded) by using data from the BSE SENSEX. Existence of several stylized facts are corroborated as discussed in the study.
- His study is enriched by the application of several sophisticated analytical techniques including extensive usage of models from the GARCH school as well as their comparisons, the Hannan-Quinn criterion, the different tests for long memory etc.
MATERIALS AND METHODS

The study presented below considers SENSEX as a proxy for the Indian market. The period under consideration is from 1997 to March 2009. The daily return is used for all the analyses to make the series unit root free. The daily return is defined as:

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

where, \( P_t \) is the closing price on day \( t \). The analysis considers the raw return series (with sign) for the study of descriptive statistics, volatility and related topics while the absolute return series (without sign) is used for the study of long memory.

The next few sub-sections introduce the concepts and tools necessary to carry out the subsequent analysis concerning volatility and long memory and obtain the results.

Checking for unit root: Once the (raw) daily return series is constructed, the Augmented Dickey Fuller test and the Phillips Perron test are performed to check the presence of unit root.

Computation of the descriptive statistics: The basic understanding about the raw return series can be found from the descriptive statistics. Two most important statistical measures for the series are the kurtosis and the Jarque Bera statistic. Kurtosis of a series is a measure of the flatness of the distribution. The Jarque Bera test statistic shows whether the series follows the normal distribution.

Study of volatility using models from the GARCH school: To understand the embedded volatility in the series, models from the GARCH (Generalized Autoregressive Conditional Heteroskedasticity) school are employed because of their ability to capture the conditional volatility and also the asymmetry present in it, if any. It is possible to use GARCH models when the series is non-normal. The models that are used in this analysis are explained briefly.

GARCH (1,1): This is the simplest form of GARCH process. First used by Engle (1982) and later modified by Bollerslev (1986), it has a first order autoregressive GARCH term and a first order moving average ARCH term and hence the nomenclature (1,1). This model is specified by a mean equation and a variance equation.

The mean equation: \[ Y_t = X_t \theta + \epsilon_t \]  
(1)

The variance equation: \[ \sigma_t^2 = \omega + \alpha \epsilon_{t-1}^2 + \beta \sigma_{t-1}^2 \]  
(2)

Where:
- \( X_t \): Exogenous variable
- \( \sigma^2 \): The conditional variance
- \( \omega \): Constant
- \( \epsilon_{t-1}^2 \): News about volatility from the previous period, measured as the squared residual from the previous period-this is the ARCH term
- \( \sigma_{t-1}^2 \): Previous period’s forecast variance-this is the GARCH term
GARCH (q, p): This is a more generalized form and denotes higher order GARCH model with q as the GARCH term and p as the ARCH term.

The conditional variance for a GARCH (p, q) process is specified as:

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \beta_i \sigma_{t-i}^2 + \sum_{i=1}^{p} \alpha_i \epsilon_{t-i}^2$$

(3)

$\epsilon_{t-1}$ is the ARCH term, denoting news about volatility from the previous period and $\sigma_{t-1}^2$ is the GARCH term, denoting forecast variance of the previous period.

Models with asymmetric information shock: However, the above models cannot capture the phenomena of asymmetry of volatile response of the market towards positive and negative announcements (Engle and Ng, 1993). The asymmetry is characterized by the fact that an unexpected bad news or a negative shock has a greater impact on the market than the effect of a good news or positive shock of the same magnitude (gain/loss asymmetry in the language of stylized facts). The Exponential GARCH (EGARCH) and Threshold GARCH (TGARCH) models allow for such possibility.

EGARCH: The EGARCH (Exponential GARCH) model was proposed by Nelson (1991). Here the conditional variance is expressed as a logarithmic term, which shows an exponential leverage effect. If the hypothesis that $\gamma_1 < 0$ is true, a leverage effect is said to be present. Presence of asymmetry is confirmed if $\gamma_1 \neq 0$.

TARCH: The TARCH (Threshold GARCH) model is also used to study the asymmetric nature of the volatility (Zakoian, 1994; Glosten et al., 1993). The TGARCH model is often called GJR model after Glosten, Jagannathan and Runkle.

In this model, good news, $\epsilon_{t-1} > 0$ has an impact $\alpha_1$, while bad news $\epsilon_{t-1} < 0$, has an impact $\gamma_1 + \gamma_2$. If $\gamma_1 > 0$, bad news increases volatility and there exists a leverage effect. If $\gamma_1 > 0$ then the impact is asymmetric. This model takes into account a quadratic leverage effect term, rather than an exponential one.

Asymmetric component GARCH model: This model can capture both short-run and long-run volatility and also the asymmetry in the volatility. This model allows for a mean reversion to a varying level.

Study of long memory: The study now tries to unravel whether there is any long term dependence or long memory in the market. A random process is said to possess long-memory if it has an autocorrelation function which is not integrable. This implies that the autocorrelation function decays asymptotically as a function of the time lag. Such a strong autocorrelation implies a high degree of long-term predictability, meaning thereby that present stock prices can be significantly affected by prices from the distant past. If long memory is present in the system, a shock once propagated in the system continues to affect future outcomes even after significant amount of lags. This increases risk in the market and hence it has a profound effect on international trade and investment policies of different countries.

The most widely used test to investigate the nature and extent of long-term memory in a time series, is the Rescaled Range (R/S) analysis. This test was first used by Hurst (1951) to test for long
memory in the pattern of flooding by the river Nile. It was later modified by Mandelbrot (1972, 1975), Mandelbrot and Taqqu (1979) and Mandelbrot and Wallis (1968, 1969).

The basic idea underlying the classical R/S test is to compare the minimum and maximum values of running sums of deviations from the sample mean, renormalized by the sample standard deviation. For long-memory processes the deviations are larger than for processes which do not have long-memory.

Consider a time series $y_t$ for $t = 1, 2, \ldots, T$. The classical R/S statistic is defined as:

$$
Q_T = \frac{1}{S_T} \left[ \max_{i \leq T} \sum_{i=1}^{k} (y_i - \bar{y}) - \min_{i \leq T} \sum_{i=1}^{k} (y_i - \bar{y}) \right]
$$

where $y_1, \ldots, y_T$ are sample observations and $\bar{y}$ is the sample mean given by:

$$
\bar{y} = \frac{1}{T} \sum_{i=1}^{T} y_i
$$

The term within bracket shows the range of partial sum of deviations and it is rescaled by dividing with $s_T$, the standard deviation which is defined as:

$$
s_T = \left[ \frac{1}{T} \sum_{i=1}^{T} (y_i - \bar{y}) \right]^{1/2}
$$

The classical R/S test has been proven to be too weak, i.e., it tends to indicate that a time series has long memory when, it does not really have so. Lo (1991) criticized the R/S statistic for being unable to distinguish between short term and long term dependence. In order to remove this shortcoming Lo modified the classical R/S statistic in such a way that its statistical behaviour is invariant over a general class of short memory processes, but deviates for long memory ones. This is done by replacing the denominator which is the square root of a consistent estimator of the variance of the partial sum. The estimator involves not only sums of squared deviations of $y_i$, but also its weighted auto-covariances up to lag $q$.

The present study uses both these models (R/S statistic and modified R/S statistic) to test for long term memory.

The study on long memory concludes with an examination of whether the series is characterized by fractional integration. A fractionally integrated process is neither a perfectly short memory I(0) process, nor a I(1) but falls between these two processes, thus, possessing weak correlation between observations situated far apart in the series. To test for fractional integration, the GPH test is used.

RESULTS

The total sample size considered in this investigation was $T = 2777$.

The results obtained could be classified under three categories:

- Results from the unit root tests and descriptive statistics
- Results from the study of volatility
- Results from the tests for long memory
Unit root tests: The filtered raw return series (AR-GARCH filtered) were checked for possible presence of unit root. Augmented Dickey Fuller test and Philips Perron tests were carried out to check the stationarity of the raw return time series. The results were summarized in Table 1. The null and alternative hypotheses were as follows:

- H₀: Unit root is present
- H₁: Unit root is not present

The tests were carried out at 1% level of significance.

Values of t-statistic were computed and the p-values were determined. The extremely low p-values indicated that there was sufficient statistical evidence to reject the Null hypothesis. In other words, it could be inferred that the raw return series was free from unit root.

Descriptive statistics: The results for the descriptive statistics were displayed in Table 2 where values of the mean, median, maximum and minimum values, standard deviation, skewness and kurtosis were shown. The kurtosis (value > 3) for the raw return series revealed that the series had a fat tail and that it was leptokurtic (more peaked than normal). The series was also found to be negatively skewed.

Jarque-Bera test: The above test was carried out under the following Null and Alternative hypotheses:

- H₀: The series follows a normal distribution
- H₁: The series is non-normal

The computed value of the JB test statistic was 1252.12 and the p-value was found to be 0. The extremely low p-value indicated that there was sufficient statistical evidence to reject the null hypothesis and conclude that the series was non-normal.

<table>
<thead>
<tr>
<th>Table 1: Result of unit root tests</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADF Test</td>
</tr>
<tr>
<td>-----------------------------------</td>
</tr>
<tr>
<td>Results</td>
</tr>
<tr>
<td>Raw return</td>
</tr>
<tr>
<td>Null hypothesis</td>
</tr>
</tbody>
</table>

*Indicates significance at 1% level (1% critical value = -3.961)

<table>
<thead>
<tr>
<th>Table 2: Descriptive statistics for raw SENSEX daily return</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statistical measure</td>
</tr>
<tr>
<td>--------------------</td>
</tr>
<tr>
<td>Mean</td>
</tr>
<tr>
<td>Median</td>
</tr>
<tr>
<td>Maximum</td>
</tr>
<tr>
<td>Minimum</td>
</tr>
<tr>
<td>Standard deviation</td>
</tr>
<tr>
<td>Skewness</td>
</tr>
<tr>
<td>Kurtosis</td>
</tr>
</tbody>
</table>
Results from application of the GARCH models: In this analysis, first a GARCH model was employed to test for the presence of volatility. Then TARCH and EGARCH models were employed and the best fit model among the two, based on the lowest value of the selection criterion (Hannan-Quinn for present study) was analyzed to examine possible presence of asymmetry and leverage effect in the series. Lastly, the asymmetric component GARCH model was used to check for the mean reverting property of the series.

The results from the GARCH School of models were summarized below in Table 3.

The lags were chosen according to the Hannan-Quinn criterion. As the series had a large number of observations, Hannan-Quinn criterion was a better selection criterion than the Schwartz or Akaike criterion (Shittu and Asemota, 2009). From Table 3, it could be seen that news about volatility from past periods had a significant effect on the current period’s volatility. The coefficients of the GARCH term were significant at 10% level in all lags except the third. So, it could be said that in the first, second, fourth and fifth lags, the current period’s volatility was dependent on the past period’s volatility.

Among the TARCH and EGARCH models, EGARCH was the better model because the Hannan-Quinn criterion was lower for EGARCH and the term indicating asymmetry was significantly negative. Hence it could be said that the series showed significant amount of asymmetry. Finally, the component GARCH model showed that the coefficient of the term indicating long run persistence was 0.99 and therefore there existed strong volatility persistence.

Finally, the ARCH-LM test was employed to examine whether any ARCH effect remained in the residual. Ignoring the presence of ARCH effect might lead to inefficient results. However, up to 5 lags were taken and there was no evidence of any remaining ARCH effect.

<table>
<thead>
<tr>
<th>Table 3: Summary of results from the models of the GARCH School for Raw SENSEX daily return series</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable</td>
</tr>
<tr>
<td>---------------------------</td>
</tr>
<tr>
<td>GARCH</td>
</tr>
<tr>
<td>RESID(1)^2</td>
</tr>
<tr>
<td>RESID(2)^2</td>
</tr>
<tr>
<td>RESID(3)^2</td>
</tr>
<tr>
<td>GARCH(-4)</td>
</tr>
<tr>
<td>GARCH(-2)</td>
</tr>
<tr>
<td>GARCH(-3)</td>
</tr>
<tr>
<td>GARCH(-4)</td>
</tr>
<tr>
<td>GARCH(-5)</td>
</tr>
<tr>
<td>TARCH</td>
</tr>
<tr>
<td>RESID(1)^2*(RESID(-1)&lt;0)</td>
</tr>
<tr>
<td>Hannan-Quinn crit</td>
</tr>
<tr>
<td>EGARCH</td>
</tr>
<tr>
<td>O(7)</td>
</tr>
<tr>
<td>Hannan-Quinn crit</td>
</tr>
<tr>
<td>OGARCH</td>
</tr>
<tr>
<td>O(3)</td>
</tr>
</tbody>
</table>

**ARCH-LM**

<table>
<thead>
<tr>
<th>Statistical measure</th>
<th>Coefficient</th>
<th>Results</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>LAG 1</td>
<td>F-statistic</td>
<td>1.662</td>
<td>Prob. F (1.2775)</td>
</tr>
<tr>
<td></td>
<td>Obs*R-squared</td>
<td>1.562</td>
<td>Prob. Chi-Square (1)</td>
</tr>
<tr>
<td>LAG 2</td>
<td>F-statistic</td>
<td>1.514</td>
<td>Prob. F (1.2773)</td>
</tr>
<tr>
<td></td>
<td>Obs*R-squared</td>
<td>3.028</td>
<td>Prob. Chi-Square (2)</td>
</tr>
<tr>
<td>LAG 3</td>
<td>F-statistic</td>
<td>1.596</td>
<td>Prob. F (3.2771)</td>
</tr>
<tr>
<td></td>
<td>Obs*R-squared</td>
<td>4.788</td>
<td>Prob. Chi-Square (3)</td>
</tr>
<tr>
<td>LAG 4</td>
<td>F-statistic</td>
<td>1.294</td>
<td>Prob. F (4.2760)</td>
</tr>
<tr>
<td></td>
<td>Obs*R-squared</td>
<td>5.177</td>
<td>Prob. Chi-Square (4)</td>
</tr>
<tr>
<td>LAG 5</td>
<td>F-statistic</td>
<td>1.066</td>
<td>Prob. F (5.2767)</td>
</tr>
<tr>
<td></td>
<td>Obs*R-squared</td>
<td>5.334</td>
<td>Prob. Chi-Square (5)</td>
</tr>
</tbody>
</table>
When the raw returns were plotted against time, another distinct feature of the SENSEX return series could be identified as shown in Fig. 1. The series showed sustained periods of high and low volatilities. In other words, volatility clustering could be said to be present in the series.

**Results from the tests of long memory:** Absolute return was considered for the long memory tests because the raw returns tended to have a mean very close to zero and hence might indicate the absence of long memory when it was actually present.

The R/S and modified R/S tests were done using the following Null and Alternative hypotheses:

- $H_0$: There is no long term dependence in the return series
- $H_1$: There exists long term dependence in the return series

Fig. 1: Plot of raw daily return on SENSEX against time for the period extending from 1997 to 2009

Fig. 2: Plot of Autocorrelation Function (ACF) for absolute daily return on SENSEX for the period extending from 1997 to 2009
Table 4: Results from tests of long memory and fractional integration for absolute daily return series

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>R/S test</th>
<th>Modified R/S test</th>
<th>GPH test (d)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Null Hypothesis</td>
<td>No long term dependence</td>
<td>No long term dependence</td>
<td></td>
</tr>
<tr>
<td>Absolute Return</td>
<td>4.0079**</td>
<td>7.0755**</td>
<td>0.3815*</td>
</tr>
</tbody>
</table>

*indicates significance at 1% level; GPH test critical value = 3.7885

The tests were carried out at the 1% level of significance.
The following values were obtained:

- R/S test statistic $Q_t = 4.0079$ and the modified R/S test statistic $Q_t = 7.0755$

Values in both cases were greater than the corresponding critical values at the 1% level of significance which indicated that there was sufficient statistical evidence to reject the Null hypothesis and conclude that there was significant presence of long term memory in the absolute return series.

The d value from the GPH test suggested that the series was fractionally integrated; all these results were summarized in Table 4.

The long memory property could also be illustrated with help of a diagram. Figure 2 showed the autocorrelation function upto 200 lags. As could be seen from the figure, the autocorrelations of the absolute returns were highly persistent and although slowly decaying, clearly indicated presence of long memory in the SENSEX return series.

DISCUSSION

The stylized facts occurring in financial markets have been reviewed by different authors (Cont, 2001; Coolen, 2004). Several different stylized facts like fat tails, volatility clustering, volatility persistence and long memory have been discussed (Beran, 1994; Ding et al., 1993; Greene and Fielitz, 1977). Attempts have been made to interpret the stylized facts as emergent properties of a complex system, viz., the financial market (Rickles, 2008). Stylized facts including the leverage effect have been discussed by Bouchaud and Potters (2001), while they have been related to minority games by Challet et al. (2001). The dynamics of financial markets including stylized facts have been discussed by Reimann and Tupak (2010).

Long term dependence (or long memory) has also been a topic of intense research. Greene and Fielitz (1977) took 200 daily stock return series and used the R/S analysis to show the presence of long term memory in many of them. They concluded that in the presence of long term dependence, martingale process does not hold. Ding et al. (1993) discovered long memory stochastic volatility in stock returns. Baillie et al. (1996) in their seminal paper introduced the Fractionally Integrated Exponential GARCH model (FIEGARCH) to capture the long run dependence in the US stock market.

The objective of the present study was to highlight the importance of the stylized facts and provide empirical evidence for their existence by using data from the Indian stock market and subjecting them to statistical analysis. The raw daily return was found to be a non-normal distribution having fat tail. Volatility clustering, volatility persistence as well as asymmetry in volatility were all observed in the raw return. Impact of past news was also observed. Further, the absolute return series displayed long memory features.
Thus, the foregoing analysis showed ample justification for the existence of the stylized facts. The stylized facts are intimately related to the complexity argument; hence justification of these facts indicates that a financial market may be regarded as a complex system.

ACKNOWLEDGMENTS

The authors acknowledge the support of West Bengal University of Technology (WBUT) under Grant at WBUT/UGC/F.5-216/2006 (HRP), January 2007.

REFERENCES