Stylized facts of the Indian Stock Market

Rituparna Sen

Indian Statistical Institute, Bangalore

01.10.2019

Joint work with Manavathi S.

Table of contents

Introduction

- Any stock market deals with the trading of shares of stocks which involves inherent uncertainty.
- Because of the huge amounts of money involved, economists use different tools and measures to develop models for stock prices.
- These models are then used for portfolio selection, risk management, derivative pricing etc.

Introduction

- Stylized facts are properties that are common across various markets and time domains. These properties offer a way to generalize stock price behavior irrespective of the instruments used.
- It is reasonable to expect the prices arising from any acceptable model to follow at least these properties.
- Lists of several such stylized facts are available in the literature for the developed western markets.
- In [3] we explore these properties for the Indian market.

Description of data

- The historical data is downloaded from yahoo finance.
- Fifty stocks that are traded in the National Stock Exchange(NSE) are considered. The stocks are the constituents of the NIFTY index.
- For each stock the data contains seven attributes for each trading day, namely, the date, closing price, opening price, the highest price, lowest price, volume traded and the change percentage on that day.
- These were taken over the time range of January 2007 to November 2017.

Date	Close	Open	High	Low	Vol.	Change %
Nov 30, 2017	400.4	399.25	401.3	397.1	611.03K	-0.76
Nov 29, 2017	403.45	397.6	405.6	395.05	5.21M	1.64
Nov 28, 2017	396.95	399.6	403.65	395.2	6.01M	-0.64
Nov 27, 2017	399.5	404.5	404.5	398	2.99M	-1.26
Nov 24, 2017	404.6	411	411	401.85	1.37M	-0.61
Nov 23, 2017	407.1	415.3	415.3	402.55	3.90M	-1.79
Nov 22, 2017	414.5	402.5	416.65	401.65	5.17M	3.42
Nov 21, 2017	400.8	399.3	405.85	399.3	5.42M	0.55
Nov 20, 2017	398.6	400.95	405.9	397.1	4.44M	-0.59
Nov 17, 2017	400.95	401	405	396	6.52M	0.94
Nov 16, 2017	397.2	408	410.6	395	5.32M	-2.43
Nov 15, 2017	407.1	410	414.25	403.65	1.85M	-0.89
Nov 14, 2017	410.75	413.35	417.85	406.7	2.33M	-0.63

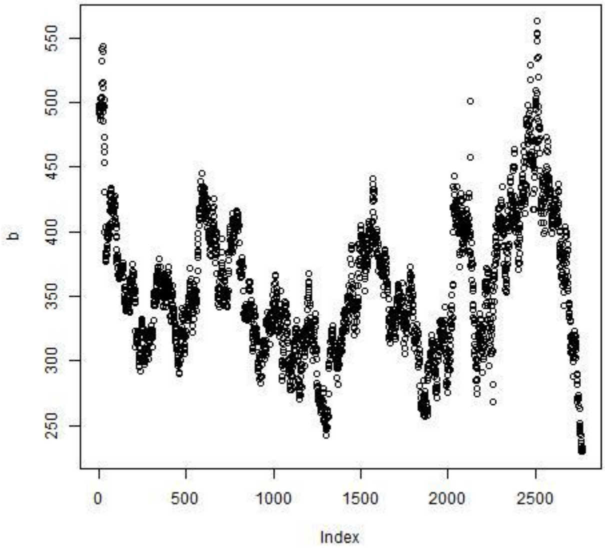


Figure 1: Prices of Bosch

- Since the prices are often **non-stationary**, it is more common to use log returns for statistical analysis.
- The log return of a stock $R(t)$ at time t is given by

$$R(t) = \log(S(t)) - \log(S(t-1))$$

  where $S(t)$ denotes price of stock at time t.
- For the rest of the paper we deal with the log return and not the prices.
- For illustrative purpose, the log returns for the Bharat Petroleum Corporation Limited (BPCL) stock are shown in Figure 2.
- A visual comparison with Figure 1 shows that the concern of non-stationarity has been addressed.

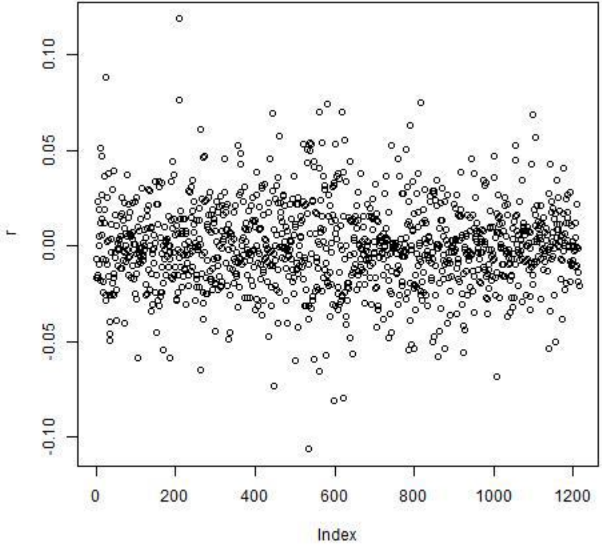


Figure 2: Log Returns of BPCL

Simple Distributional properties

- First we explore the simple stylized facts related to the distribution of returns.
- We ignore time series dependence and tail behaviour.
- In particular, we study the symmetry and normality of the distribution, as well as the inter-dependence of the return and volatility.

Gain loss asymmetry

- A common stylized fact is the gain loss asymmetry. One observes large drawdowns in stock prices and stock index values but not equally large upward movements.
- The skewness of a random variable X is the third standardized moment and is denoted by γ_1.

$$\gamma_1 = \frac{E(X-\mu)^3}{(\sqrt{E(X-\mu)^2})^{\frac{3}{2}}},$$

  where $\mu = E(X)$ is the mean of the distribution.
- Skewness measures the asymmetry of the probability distribution of a random variable.

Introduction

- A positive skew distribution means that the right tail is longer than the left tail. A negative skew distribution means that the left tail is longer.
- Large drawdowns compared to upward movements would correspond to a long left tail and hence, a **negatively skewed curve**.
- For the data set under consideration, skewness of the returns was calculated.
- Most stocks have **positive skewness** as shown in Figure 3.

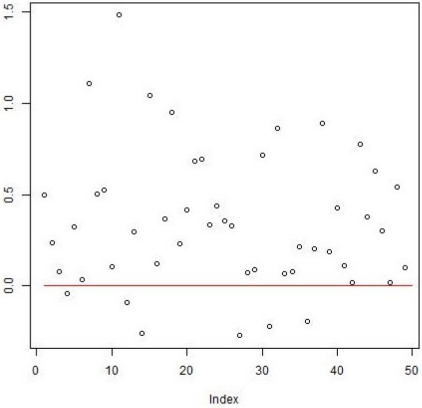
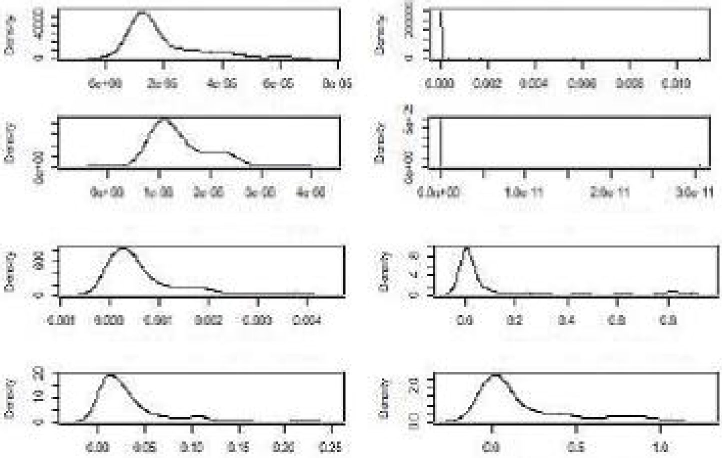


Figure 3: A plot of the skewness values.

- It is observed that most stocks have positive skewness and thus show larger upward movements than drawdowns. This is in contrast with the stylized facts reported in [1].
- In [2], the authors compare asymmetry indices of historical prices from ten stock markets using market index data. They find that in most stock markets, price fall is faster than price rise; while in China and India, price rise is generally faster than price fall.
- Our results reconfirm this for the Indian market at the stock level.

Aggregational Gaussianity

- The next stylized fact is aggregated Gaussianity, which is the following phenomenon. As one increases the time scale Δt over which returns are calculated, their distribution looks more and more like a normal distribution.
- In particular, the shape of the distribution is not the same at different time scales.
- Two normality tests in R are performed on the data, namely the Kolmogorov-Smirnov (KS) test and the Shapiro-Wilke (SW) test for daily, weekly, monthly and quarterly returns.
- The p-values increase as the time over which the returns are calculated increases. The kernel density plots of these p-values for the 50 stocks under consideration are presented in Figure 4.


Density
0.000 0.002 0.004 0.006 0.008 0.010
-0.001 0.000 0.001 0.002 0.003 0.004
0.0 0.2 0.4 0.6 0.8
0.00 0.05 0.10 0.15 0.20 0.25
0.0 0.5 1.0

Aggregational Gaussianity

- It is seen that as the time scale increases, the p-vales are less and less concentrated around zero and the normality assumption is rejected for fewer stocks.
- Thus, the distribution of returns of many stocks become similar to the normal distribution as the time over which returns are calculated is lengthened. We conclude that aggregated Gaussianity is present in the data.

Leverage Effect

- Leverage effect refers to the observation that most measures of volatility of an asset are negatively correlated with the returns of that asset.
- If there is high volatility in the stock movement, then the returns will be low.
- The rationale behind this is that if there is a lot of fluctuating movement of the stock price, not many investors will invest in the stock.
- The correlation between returns and squared returns is calculated. Figure 5 shows a density plot (for all fifty stocks) of the correlation between volatility and returns.
- It is expected to be **negative**, but is found to be **positive**. Thus the observed results are contrary to the fact stated in [1].

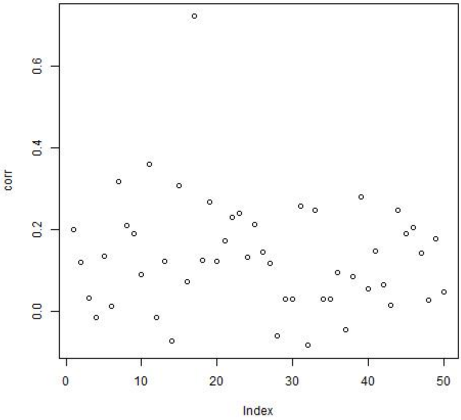


Figure 5: Correlation between returns and volatility.

Time Series Dependence

- The autocorrelation of a stationary time series $X_t, t = 1, \cdots, T$, denoted by ρ_X is defined as a function of the lag as.

$$\rho_X(k) = cor(X_s, X_{s+k}) = \frac{E(X_s - \mu_X)(X_{s+k} - \mu_X)}{\sigma_X^2}.$$

  Here μ_X and σ_X^2 are respectively the mean and variance of X_s. Due to stationarity, μ_X, σ_X^2 and $\rho_X(k)$ do not depend on s.

- The autocorrelation of a time series measures the linear correlation between the original time series (X_s) and the lagged series (X_{s+k}).

Autocorrelation of returns

- The autocorrelation depends on the lag k. For example, the autocorrelation of a time series X with the daily returns of a stock with lag 1 would indicate how much a day's return influences the next day's return, or how much a day's return depends on the previous day's return.
- Partial autocorrelation $\phi_X(k)$ measures the influence of the value at a given instant on the value at the instant at lag k, controlling for the effect of intermediate values. That is,

$$\phi_X(k) = cor(X_s, X_{s+k} | X_{s+1}, \cdots, X_{s+k-1}).$$

Autocorrelation of returns

- The stylized fact is that autocorrelations of asset returns are often insignificant, except for very small intraday time scales for which microstructure effects come into play.
- This fact is the reason why investing in stocks is risky. This is why it is difficult to predict the future stock prices.
- If the returns were correlated, then it would mean that the return values are dependent on previous return values, and the correlation coefficient can be used to determine the expected value of the future return and hence price.

Autocorrelation of returns

- Here the autocorrelation and partial autocorrelation for all the 50 series has been calculated for the returns of the stocks using a lag of 10.
- These quantities are estimated using the method of moments estimators.
- For demonstrative purpose, Figure 6 shows the partial autocorrelation (PACF) values for the Maruti Suzuki India Ltd (MRTI) stock. The dashed lines mark the 5% significance levels. Observe that all partial autocorrelation upto order 10 are insignificant.

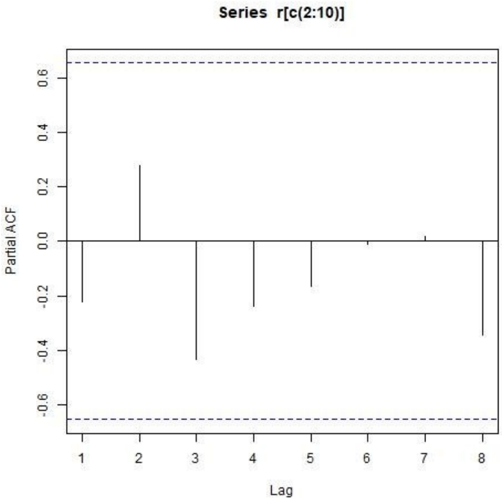


Figure 6: Partial autocorrelation values for MRTI.

- For a formal evaluation, the Portmanteau tests are carried out to test the hypothesis that all autocorrelations upto a certain lag are zero. These tests are to determine lack of fit of the data. That is, they determine how close the data are to white noise.
- This gives a measure of how much one value depends on the previous value(s).
- There are two Portmanteau tests: Box-Pierce and Ljung-Box. Both tests give similar results. So we present the results for the Ljung-Box test only.

- The Ljung-Box test statistic is given by

$$Q = n(n+2)\sum_{k=1}^{m}\frac{\hat{\rho}_k^2}{n-k} \tag{1}$$

- Here this variable Q is follows a χ^2 distribution with m degrees of freedom under the null hypothesis. Here n is the total number of observations, m is the maximum lag up to which autocorrelation is determined.
- Using a maximum lag of 10 and level of significance 1%, the null hypothesis is rejected for 22 stocks out of 50.
- This implies that there is some autocorrelation in many of the stocks under consideration. This is again at variance with the observations from other markets and can potentially be useful to predict future prices, giving rise to arbitrage opportunities.

Volatility Clustering

- Different measures of volatility display a positive autocorrelation over several days, which quantifies the stylized fact that high-volatility events tend to cluster in time.
- If there is volatility clustering, it means that there is some significant autocorrelation in the volatility of the stocks, so that the values depend on the previous values and tend to be similar over time.
- So if the volatility on one day is high, then that will cause the volatility to remain high over the subsequent days.

Volatility Clustering

- One measure of the volatility of a stock is the variance of the returns.
- On calculating the average return (μ) for all the stocks, most stocks had average return close to zero.
- So for this data set the squared returns can be considered as a fairly accurate measure of volatility.
- The autocorrelation of squared returns was calculated for a lag upto 10 and formal hypothesis tests are conducted using absence of autocorrelations between squared returns as null hypothesis and test statistic: $X = \frac{\sqrt{n}\hat{\rho}}{1-\hat{\rho}^2}$ where n is the number of observations, and $\hat{\rho}$ is the estimated autocorrelation.

Volatility Clustering

- Under the null hypothesis, X follows a standard normal distribution asymptotically. This can be used to compute the p-value of the test.
- Majority of p-values were less than 5%. The plots of p-values of the squared returns of HDFC Bank (HDFC) stock is shown in Figure 7.
- Portmanteau tests were also conducted. All the stocks show a non-zero autocorrelation between the squared returns. The autocorrelation at lag one is positive in almost all cases.

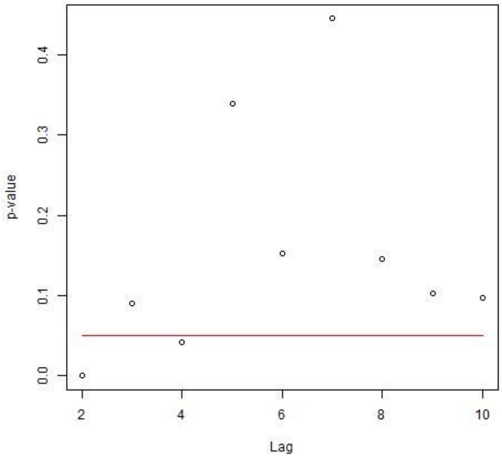


Figure 7: p-values for ACF of squared returns of HDFC.

In summary, the data does display volatility clustering, as expected.

Heavy Tails

- The stylized fact regarding tails states that the distribution of returns seems to display a power-law or Pareto-like tail, with a tail index which is finite, higher than two and less than five.
- A power law is a functional relationship between two variables, where the relative change in one variable is proportional to the relative change in the other.
- It is of the form
$$Y = kX^{\alpha}$$
where X and Y are variables of interest, α is the law's exponent and k is a constant.

Heavy Tails

- Distributions of random variables are studied in comparison with the exponential distribution.
- The tail of a distribution is the part of the distribution where $|X|$ tends to ∞. The thickness of the tail is the tail index.
- Distributions can be classified as being heavy tailed or light tailed. A heavy-tailed distribution has a tail that is not bounded by the exponential tail, whereas the light-tailed distribution has a tail that falls below the exponential tail.
- Here we look only at the tail and not the part of the distribution before where the tail begins.
- The choice of the point in the distribution where the tail begins is also important in determining tail index.

Heavy Tails

- Consider any distribution $P(X)$ with cumulative distribution function $F(x) = 1 - \overline{F}(x)$ defined by $Pr(X > x) = \overline{F}(x)$, such that for some ξ ¿0,

$$\overline{F}(x) = x^{\frac{-1}{\xi}} L(x)$$

  where $L(x)$ is some slowly varying function for large x.
- The tail index of the fat-tailed distribution $P(X)$ is by definition ξ.
- Using the hill.adapt() function in the extremefit package of R, the tail index is calculated. The values for all 50 stocks are shown in Figure 8. It is observed that the returns indeed have heavy tails and the index is found to lie between two and five.

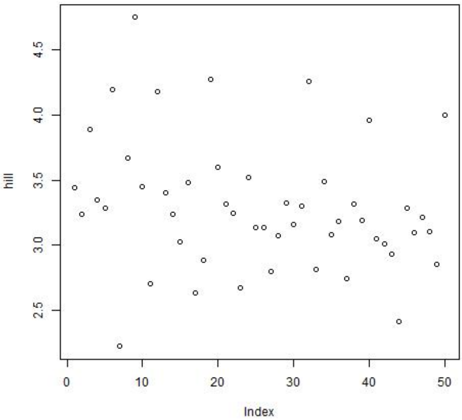


Figure 8: Tail Index of returns.

Conclusion

- Stylized facts for a large number of stocks of the Indian stock market were studied. After a brief survey of the literature on stylized facts, the data is described. Then, the facts were analyzed using the basic functions and toolboxes provided in R. Based on the observations made, certain inferences are drawn and stated.
- Among the accepted stylized facts, the ones that are found in the Indian market are aggregated Gaussianity, volatility clustering, heavy tails with an index between 2 and 5. In particular this excludes stable laws with infinite variance and the normal distribution. However the precise form of the tails is difficult to determine.

Conclusions

- There are some significant deviations of the Indian data from the stylized facts listed for developed markets.
- The gain loss asymmetry in reversed, that is, most stocks show larger upward movements than drawdowns.
- The leverage effect is also reversed for most stocks, that is, the returns and volatility are positively correlated.
- 22 out of 50 stocks show significant autocorrelation in the returns. The tail-index is not reduced by GARCH-fitting in most cases.
- This analysis can be used to decide what to test for in a predictor model for the Indian data.
- New investment techniques can be devised using the leverage, asymmetry and autocorrelation, that are different from those used in other markets.

References

[1] Rama Cont. Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance*, 1:223–236, 2001.

[2] W. Huang and Y. Zhang. Asymmetry index of stock price fluctuations. *Journal of Global Economics*, 2014.

[3] Rituparna Sen and Manavathi Subramaniam. Stylized facts of the indian stock market. *Asia-Pacific Financial Markets*, pages 1–15, 2019.

Thank You