

STATISTICAL ANALYSIS OF INDIAN STOCKS

IEOR 4150



Submitted to:

Submitted by:

Prof. Antonius Dieker

Ayush Goyal - ag3898

Bhargava Kumar - bk2668

Divas Verma - dv2394

Introduction

We have selected 10 stocks that are listed on Nifty 50 index of Indian Stock Market on the basis of the trading volume and the size of the organization. We have selected a good mix of public issued private and government run organizations.

From this project, we plan to formulate investment strategies by deriving meaningful insights from statistical tests on the selected stocks.

Analysis of the stocks include calculating daily log-returns of each individual stock and later comparing log-returns of pairs of stocks to determine correlation between them.

Data Set Description

Our data set comprises information of stock prices of 10 Indian conglomerates listed on the Nifty 50 index of National Stock Exchange (NSE), information including daily closing and opening stock price, high and low prices of the day for each stock. The above information was collected from the website of NSE of India. The scrips of the stocks and the companies they represent are listed below:

1. TCS – Tata Consultancy Services
2. SBIN – State Bank of India
3. RELIANCE – Reliance Industries Limited
4. RANBAXY – Ranbaxy Pharmaceuticals
5. ONGC – Oil and Natural Gas Corporations
6. MARUTI – Maruti Suzuki India
7. INFY – Infosys
8. HDFC – HDFC Bank
9. GAIL – Gas Authority of India Limited
10. ITC – Indian Tobacco Company Limited

These companies draw the maximum fundamental stock manipulation and are large enough to make a substantial impact on country's economy.

Goals / Objectives

We selected large scale and fundamentally driven companies form various industrial sectors, including, but not limited to, Banking, Tech, Infrastructure, Automobile and Pharma, of Indian Market.

Sectors were chosen on the basis of recent government and institutional reforms that have taken place and which have impacted the day to day activities of common people and all directly and indirectly related medium and small-scale companies.

Companies such as SIBN & HDFC were chosen because of implementation of major banking reforms. Similarly, INFY & TCS were chosen due to change in public sentiment in the U.S.A towards outsourcing of tech work and immigration policy. Due to massive national and international news coverage, we believe that these stocks show potential investment opportunity for public.

Our objective is to conduct multiple statistical tests on the daily returns of one year of these stocks to devise investment strategies that will help people to make more informed decisions.

Preparatory Work

To run statistical tests on the stocks, we required certain information with relation to the values of log returns of the stock prices. The information was captured through the following steps listed herein below:

- By finding out the consistency of log-return values with a random sample; and
- By determining the distribution of the log returns.

Using such key information, we wanted to, by means of statistical tests, evaluate the volatility of each of the stocks and try to identify any trends in the respective sectors. Such trends, if existed, either negative or positive, could help people understand the geo-political impact on markets and also highlight benefits of developmental reforms over populist reforms.

Analysis – Single Stock

Non-randomness Determination:

To determine the non-randomness of the values of log-returns, i.e. whether the stock log-returns are consistent with a random sample, we used runs test. The null hypothesis defines that the values in the data set are mutually independent which means that the data is consistent with the random sample. The alternative hypothesis is that the data-points/values in the data set are not mutually independent and therefore the data is inconsistent with the random sample.

Runs Test for Determining Non-Randomness	
Stock	P-Value
TCS	0.4488
SBIN	0.4488
RELIANCE	0.4488
RANBAXY	0.2068
ONGC	0.6136
MARUTI	0.1649
INFY	0.1298
HDFC	0.0387
GAIL	0.1298
ITC	0.3769

From the runs test, it was found that the P-Value of all the stocks is greater than 0.05, except for the HDFC stock which is equal to 0.0387. Therefore, we fail to reject the null hypothesis for stocks including, SBIN, TCS, RELIANCE, RANBAXY, ONGC, MARUTI, INFY, GAIL, and ITC. However, we can reject the null hypothesis for HDFC stock and thus conclude that the log- returns of the HDFC stock is inconsistent with a random sample and log-returns of the rest of the stocks in consistent with a random sample.

Histograms:

To test the distribution of log returns of each of the stocks we plotted their respective histograms which illustrated the frequency distribution of the stocks. Histograms of all the stocks showcased an approximate normal distribution curve for the log-returns, however one cannot say this with certainty just on the basis of histogram. To support the illustrations of histogram, a normal probability plot, also known as the Q-Q plot, and density plots for each of the stocks were produced. The Q-Q and density plots hinted at the possibility that the log-returns of some of the stocks may come from a distribution slightly different than a normal distribution. Having said that, the Q-Q plots did not show any major deviations in the log-return values from linearity and thus we convincingly assume that the log-return values of each of the stocks that we have selected have a normal distribution.

Statistical Tests:

After proving the normality of the log returns, we proceeded with statistical testing. Firstly, we created confidence intervals of the mean, μ , and variance σ^2 , of the log-returns for each of the stocks. Due to unavailability of the values of variances of the normal distributions from which the stock log-returns came from, we used the sample standard deviation and utilized the t-distribution with “N – 1” degrees of freedom in order to find the confidence interval for μ . Similarly, to find the confidence interval for σ^2 , we utilized a chi-squared distribution with “N – 1 degrees of freedom”.

95% Confidence Interval for μ and σ^2		
Stock	Confidence Level for μ	Confidence Level for σ^2
GAIL	[-0.005235 , 0.000423]	[0.000442 , 0.000627]
INFY	[-0.002106 , 0.001936]	[0.000225 , 0.00032]
ITC	[-0.003399 , 0.001441]	[0.000323 , 0.000459]
HDFC	[-0.001955 , 0.00284]	[0.000317 , 0.00045]
MARUTI	[-0.003781 , 0.001205]	[0.000343 , 0.000487]
ONGC	[-0.00369 , 0.000944]	[0.000296 , 0.00042]
RANBAXY	[-0.005951 , -0.00077]	[0.00037 , 0.000526]
RELIANCE	[-0.003164 , 0.00185]	[0.000347 , 0.000492]
SBI	[-0.002924 , 0.002327]	[0.00038 , 0.00054]
TCS	[-0.003643 , 0.000834]	[0.000277 , 0.000392]

The values of the 95% confidence interval for the mean, μ , were found to be small which illustrated that the mean log-return of the selected stocks were approximately zero with the range of upper and lower bounds not exceeding $\pm 1\%$. Furthermore, it was observed that the values of the 95% confidence interval for the variance σ^2 , were extremely small, suggesting that the variability in the log-returns of all of the stocks were almost negligible. Further analysis was done by developing regression models for log-returns of each of the stock against time.

Linear Regression of Log>Returns Against Time			
Stock	β_0	β_1	R^2
TCS	-0.00496665	2.805e-05	0.01288869
SBIN	-0.00188921	1.252e-05	0.00186781
RELIANCE	-0.00416099	2.759e-05	0.0099417
RANBAXY	-0.00334649	-1.1e-07	1.5e-07
ONGC	-0.00542306	3.189e-05	0.01555324
MARUTI	-0.0045517	2.57e-05	0.00872664
INFY	-0.00269074	2.052e-05	0.00846138
HDFC	-0.00199736	1.921e-05	0.0052718
GAIL	-0.00675214	3.422e-05	0.01201614
ITC	0.00066851	-1.297e-05	0.00235851

After running linear regression models on the log-return values of each of the selected stock against the “Time” attribute, we found the “ R^2 ” values for each of the model to be very close to “0” and therefore can be considered as negligible. Since, we already know that the lesser the R value, the less correlation exists between the variables. In this case, the low R values signify that there is no correlation between the log-return values of the stocks with respect to the “Time” variable. This conclusion is further supported by the scatter plots of log-returns vs. time which depicts the lack of a correlation between the two parameters. We have also provided the residual plots which do not demonstrate any trend or pattern and therefore we can say that linear regression models of log-return values with respect to “Time” cannot be used as an indicator to estimate future log-return values of these stocks.

Analysis – Two Stock

Comparison of log-return values for two different stocks out of the selected stocks were made to run further statistical tests. We used confidence level of 95% and the “two sample t-test method” to compare the two stocks. We assumed the variances, which were unknown, to be unequal to

determine whether the log-return values for the stocks were comparatively equivalent or not. Later, we subjected linear regression model on each pair of the stocks to determine correlation or similarities in the values of log-returns between them. Herein below we have only shown analysis of the stock SBIN vs the rest of the stocks for the sake of brevity of the report.

SBIN Two Sample T-Test and Linear Regression			
Stock	P-Value	Slope	R ²
TCS	0.46556	0.301802	0.066202
RELIANCE	0.80826	0.376774	0.129427
RANBAXY	0.08201	0.125084	0.015231
ONGC	0.48945	0.271954	0.057585
MARUTI	0.49894	0.388565	0.136059
INFY	0.87859	0.417084	0.103061
HDFC	0.64583	0.224285	0.041937
GAIL	0.21358	0.238876	0.066233
ITC	0.66361	0.280337	0.066767

In order to determine the correlation of the returns between the stock price of two companies, a linear regression was run between the log-return values of the two stocks to compute the slope and R² value for each pair of stock. Understanding behind values of Slope and R²:

- if the slope is closer to 1, then the returns of the base stock is highly proportional to the returns of the stock against which the base stock is regressed, i.e., the returns of both the stocks are approximately similar; and
- the higher the R² value, more significant is the correlation between the two stocks. Thus, if an R² value is closer to 1, we can determine that the two stocks are highly correlated and change in any one of them has a significant impact on the other, wherein the impact can be either positive or negative.

Analysis for Confidence Level 90%:

By the means of the “two sample t-test method” with confidence level 90%, considering “SBIN” as base stock against the rest of the selected stocks, we know that we cannot reject the null hypothesis for the stocks with P-Value more than 0.1. Therefore, from the table provided above we fail to reject the null hypothesis of the following stocks- “TCS, RELIANCE, MARUTI, INFY, HDFC, GAIL, ONGC and ITC, because they have comparatively higher P-Values which signify that they will have similar log returns. We can reject the null hypothesis for “RANBAXY”, since the P-Value is less than 0.1, and therefore conclude that the returns of these stocks will be different from that of “SBIN”.

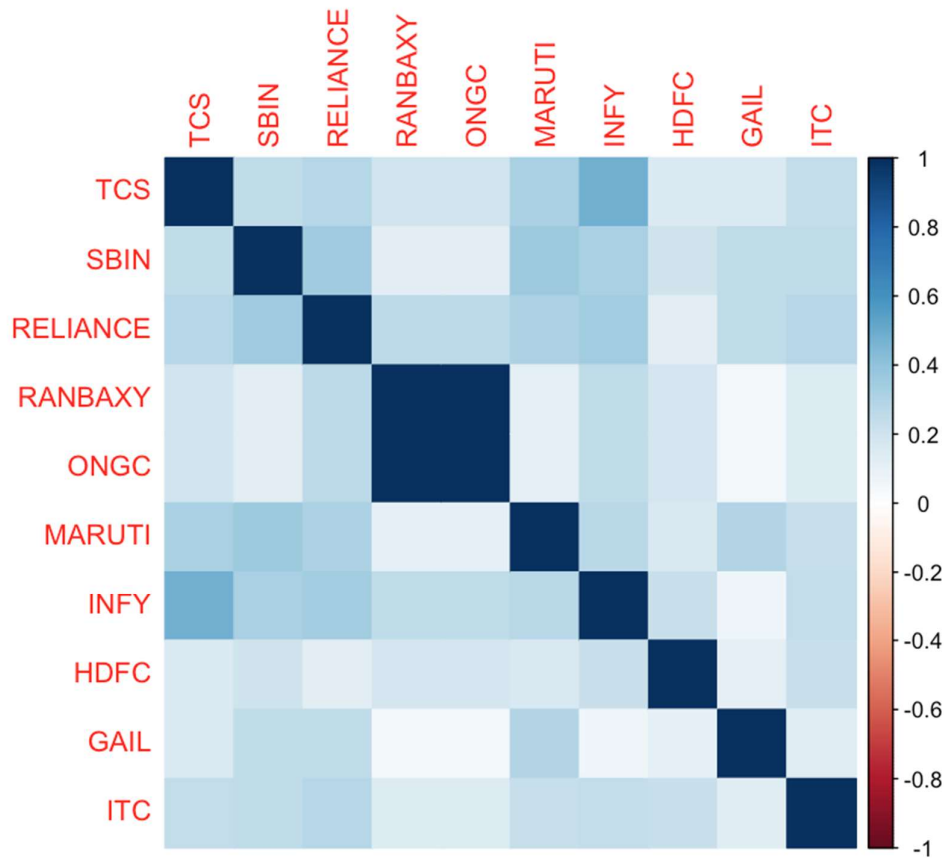
This result can somewhat be supported by the fundamental analysis of these stocks. Since, companies like TCS, RELIANCE, GAIL, and ITC have massive loan exposure to SBIN (national bank), any macro-economic bank reforms will have a similar impact on the returns of these stocks as compared to that of SBIN. We also make an interesting observation between SBI and HDFC. These two companies are the biggest public and private lending banks respectively, and should have a significant relationship with each other. We can support this statement by looking at the slope value of HDFC when regressed against SBIN. The intercept value in this case is positive for HDFC against SBIN. Whereas, in the case of “RANBAXY”, not only the P-Value is lower than “0.1”, but the R^2 and intercept values are comparatively insignificant which supports the rejection of null hypothesis for these cases and explains that we may get different returns for the amount we have invested in “RANBAXY” as compared to what we will get when similar amount is invested in “SBIN”.

Similarly, for many different pairs of stocks, such as “RANBAXY & RELIANCE”, “TCS & GAIL”, we got either the p-value less than 0.1 or R^2 value significantly low, which lead us to reject the null hypothesis and conclude that the respective stocks may have different log returns because they are uncorrelated.

Such observations were directly or indirectly supported by government actions, macro-economic reforms, banking reforms, ease of doing business reforms, geo-political deals, land reforms and many other big scale decisions taken by people of authority.

Through collective observations from our analysis we find that stocks that have low P-values tend to have low R^2 values and slopes far off from “1”, and stocks with higher P-values tend to have higher R^2 values and slopes that are comparatively closer to one. These observations also suggest that for highly correlated stocks a stronger linear relationship exists and they generally have similar returns.

Find below a pictorial correlation matrix between every possible pair of stocks.



Correlation - Sectoral Indexes vs Nifty

After analyzing “single stock” and “pair of stocks with respect to each other”, we decided to see how the Sectoral Indexes from which we had selected our individual stocks perform against the Nifty Index. The Sectoral Indexes considered during our analysis are: IT, ENERGY, FINANCE, FMCG (Consumer Goods), BANK, and AUTO.

The correlation coefficient between NIFTY and Sectoral Indexes are listed below:

NIFTY vs Sectoral Indexes	
Sectoral Index	Correlation Coefficient
IT	0.3462
ENERGY	0.8192

FINANCE	0.8829
FMCG	0.6769
BANK	0.8558
AUTO	0.7510

From the above table we can see that we get high correlation coefficients for “FINANCE”, “BANK”, and “ENERGY” Index against NIFTY, which teaches us that any changes in the return values of these Indexes will have a significant impact on the log return values of the NIFTY Index, which may drive the small scale and medium scale stocks in similar indexes.

These observations are well supported by government decisions of introducing major reforms in tax structure which impacted the “Bank Index”, and massive increase in public spending in energy sector by providing special economic zones for companies who are willing to invest in renewable energy which might have impacted the “ENERGY” Index.

Conclusion

The following conclusions were made after statistical analysis of large scale Indian Stocks from different sectoral indexes:

- Majority of our stocks have their log return values consistent with a random sample.
- Most of the selected stocks illustrate a normally distribution and therefore requires further statistical analysis.
- The linear regression of each stock against time shows that it is difficult to make any prediction on the movement of stocks.
- The correlation between different pair of stocks were determined, however none of the pair of stocks showed a highly significant correlation with each other as low R^2 values were observed.
- We could identify using sector indexes, which sector had high positive correlation on the NIFTY. If the NIFTY returns are good, then investing in such sectors can be profitable. It also helps us to build an optimized portfolio of stocks by diversifying risks.

Reference

- Call2trade – Indian stock Market Analysis software
- <https://en.wikipedia.org/wiki/Q%E2%80%93plot>
- https://en.wikipedia.org/wiki/Simple_linear_regression