**Indian Stock Markets Data: Analysis of indicators and Market Sentiments**

**Statistical Analysis Plan**

**Group Members:**

* Sagnik Ghosal
* Raman SV
* Arjun Sharma
* Adithyaa V

**Background and Rationale of Study**

Stock markets are a readily available source of real-world, structured data. They give us insight on the performance of individual companies as well as entire industries relative to the market. The ability to forecast and accurately predict the future performance of various stocks and indices on the market is paramount for an investor – newcomer and veterans alike, so as to minimize investment risk and uncertainty. Movements in stock markets are influenced by a multitude of factors such as international events, human behavior, etc. The volatility thus associated with stock performance on the market makes it challenging to accurately forecast or predict the direction of the market.

There are conflicting ideas pertaining to the predictability of the stock market. One, which advocates fundamental and technical analysis to use past data to forecast and predict stock performance, and the *Random Walk Theory* which argues that future stock prices cannot be modeled on the basis of past performance.

After studying these different schools of thought and relevant research pertaining to them, we wish to analyze whether there indeed are any particular set of methods that can reliably predict stock market performance reliably.

**Objectives and Hypotheses**

We pursue this study with the following objectives:

1. Identify at least 2 parameters (combination of indicators) that have had success on net positive trades over the past 15 years of data.
2. Look at major movements for larger debt and capital market indices to compare their returns with that of the parameters identified in step 1.

We aim to achieve the objectives outlined above by answering the questions below:

1. Are prominent market indicators reliable as to when one should buy or sell shares? In particular, we will focus on the following indicators:

* CCI (Commodity Channel Index)
* RSI (Relative Strength Index)
* PE Ratio (Price-to-Earnings Ratio)
* MACD (Moving Average Convergence Divergence)
* Ichimoku Cloud

1. Is it a better strategy to buy and hold an index compared to buying individual shares? We analyze the risk-free returns and compare them to the market returns.

**Study Overview**

**Data Description:** The data used for this study is a collection of daily reports from the Indian Stock Exchange – National Stock Exchange. The data is collected for the time period ranging from January 3, 2000 to December 30, 2022; giving us data gathered over 22 years to analyze. The data consists of daily summaries of the stocks’ price movements. [1]

There are 3 datatypes per file for the relevant fields – Text (stock name), decimal (stock prices and volumes), and date (timestamp for these values). Each file consists of the following columns:

1. **Symbol:** An entity’s identification as it is registered on the NSE
2. **Open:** The first recorded stock/index price for an entity on a given day
3. **High:** The highest recorded stock/index price for an entity on a given day
4. **Low:** The lowest recorded stock/index price for an entity on a given day
5. **Close:** The 30-minute moving average stock/index price during the last 30 minutes of trading for an entity on a given day [2]
6. **Last**: The last recorded stock/index price for an entity on a given day
7. PrevClose: The last recorded stock/index price for an entity for the preceding day of a given day
8. **TOTTRDQTY:** The total traded quantity of a particular stock on a given day
9. **TOTTRDVAL:** The total traded value of a particular stock on a given day
10. **Timestamp:** The date for which the above data has been observed.

Based on the above data from the NSE, we are also using the data from the NIFTY 50 index. The NIFTY 50 is an index of the country’s top 50 companies by market capitalization that are listed on the National Stock Exchange (NSE). From the perspective of our requirements, the data from the NIFTY 50 index is similarly structured to the NSE data for individual stocks.

**Source:** The National Stock Exchange is one of the largest stock exchanges in the world. It was established in 1992, with the vision to facilitate transparency in the Indian equity markets. Instead of trading memberships being confined to a group of brokers, NSE ensured that anyone who was qualified, experienced, and met the minimum financial requirements was allowed to trade [1].

**Data Collection Methods:** The National Stock Exchange maintains daily and monthly archives of the stock price movements. The said archive is available [here](https://www1.nseindia.com/products/content/equities/equities/archieve_eq.htm). Our method of collection here is to scrape the data from the archive page for the 22-year time period mentioned above. This is achieved through a script written in Python that would automate this process of data collection.

We procured the NIFTY 50 index’s historical data from the daily archives which are available [here](https://www.investing.com/indices/s-p-cnx-nifty-historical-data). Our method of collection here was rather simple in that the data is readily available for download as a comma separated values file (.csv) for the time period specified.

**Study Design:** The data used for this study is based on the results of events that take place in the real-world, based on which, we aim to forecast the values in the future. The nature of our study necessitates that we work with observational data.

**Statistical Analysis**

**Question 1:** Are prominent market indicators reliable as to when one should buy or sell shares?

We answer this question through multiple stages of analysis.

Assumptions:

1. The returns are normally distributed
2. The indicators used by us (CCI, RSI, P/E Ratio, MACD, Ichimoku Cloud) are independent.
3. The stock/index prices and returns at any point in time are independent.

Variables to be calculated (Variables 1-5 are elaborated on in the appendix):

1. Commodity Channel Index (CCI)
2. Relative Strength Index (RSI)
3. P/E Ratio
4. Moving Average Convergence-Divergence (MACD)
5. Ichimoku Cloud
6. Returns: The difference between individual share price on 2 different dates. For example, if shares of Tesla stock were purchased on January 5 2023, at a price of $110.5 per share, and sold on February 15 2023 at a price of $124.24 per share, our return would be $13.74 in profit per share that was sold. Essentially, it is the share price on the day of purchase subtracted from the share price on the day of sale. This variable is applicable to both individual stocks as well as indexes.

**Stage 1:** **Correlation analysis**

Correlation analysis can be used to examine the relationship between a particular market indicator and stock returns. If a high correlation is found, it may suggest that the market indicator is a useful tool for predicting stock price movements. For this study, we use the Spearman’s rank correlation coefficient, which is mathematically expressed as:

⍴ = Cov(x,y)/𝜎x𝜎y

Where x is a market indicator (mentioned in appendix) and y refers to the returns for a particular stock or index, in the context of this study.

The correlation coefficient would help inform the further stages of this study. A high level of correlation would incentivize us to perform further analysis to ascertain a potential link between any indicator (or combination of indicators) we use and the stock returns. It is noted that the correlation coefficient does not account for non-linear relationships.

**Stage 2:** **Hypothesis Testing**

**One-Sided t-Test:**

At this stage, we attempt to ascertain if buying decisions in the stock market influenced by the indicators of our interest lead to higher returns than the overall market average returns. For this comparison, the one-sample t-test will be used. We compare the mean daily returns of a group of stocks selected using each indicator with the overall market average daily returns.

We formulate our Hypothesis as follows:

**H0**: The mean daily returns of a group of stocks selected using a particular market indicator are not significantly different from the overall market average daily returns.

**H1:** The mean daily returns of a group of stocks selected using a particular market indicator) are significantly different from the overall market average daily returns.

The above test is performed for each individual indicator mentioned on page 1 (objectives and hypothesis). Hence, we perform 5 such t-tests.

**Stage 3: Regression Analysis**

We perform Linear Regression with the stock price return as our dependent variable.

The predictors for the model will include:

1. Open Price
2. Close Price
3. High Price
4. Low Price
5. Indicator (One of CCI, RSI, PE Ratio, MACD, Ichimoku Cloud)
6. Stop Loss - Price when we exit the trade
7. Target Price - The target price of the stock at which we wish to exit the trade.

We formulate our hypothesis as follows:

**H0**: All ꞵi = 0 for i = 1, 2…, where ꞵ refers to a regression coefficient.

**H1:** ꞵi ≠ 0

The above test is performed for each individual indicator mentioned on page 1 (objectives and hypothesis). Hence, we develop 5 such models, one for each indicator. We also create other models using a combination of 2 or more indicators as predictors of the returns.

**Question 2**: Is it a better strategy to buy and hold an index compared to buying individual shares?

Here, we compare market returns from individual stocks with those from the Nifty 50 index.

Assumptions:

1. The returns of the stock(s) and the Index are normally distributed.
2. The returns of the stock(s) and the Index have unequal variances [10].

Variables to be calculated:

1. Returns: Described in Question 1

We formulate our hypothesis as follows:

**H0:** There is no difference between the returns obtained through an index and those obtained through individual shares.

**HA:** The returns obtained through individual stocks are higher than those obtained through an index.

We use the Unequal Variance t-test (Welch Test), to test for a difference in the stock returns from an individual stock compared to those from an index.

The following calculations are performed:

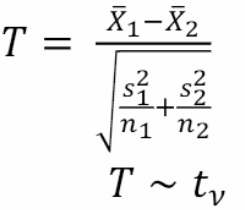
Sample Mean:

1. X̄1: Mean of stock returns.
2. X̄2: Mean of index returns.

Sample Standard Deviation:

1. s12: Standard deviation of the stock returns.
2. s22: Standard deviation of the index returns.

Test Statistic:

, 𝑇 ∼ 𝑡*v*

The p-value obtained through this test is then used to determine the veracity of the proposed hypothesis. We choose a Type-1 error probability of 0.05 (𝛼 = 0.05).

If the p-value is less than 0.05, we reject the null hypothesis, under the reasoning that there is evidence to suggest that individual stocks yield higher returns than indexes.

If the p-value is greater than 0.05, we fail to reject the null hypothesis because we do not have sufficient evidence to conclude that the returns from individual stocks are better than those obtained from the index.

**Appendix**

Table summarizing the common indicators used in our study.

|  |  |  |  |
| --- | --- | --- | --- |
| **Indicator** | **Definition** | **Mathematical Expression** | **Context of Use** |
| Commodity Channel Index (CCI) | CCI is a technical analysis indicator that measures a security's deviation from its statistical average. It is often used in the financial markets to identify overbought or oversold conditions, trend reversals, and potential buy or sell signals. CCI above +100 is considered overbought and below -100 is considered oversold. | Mean Deviation = Mean of the absolute difference between each period's TP and 20-period Simple Moving Average of TP | Helps identify potential overbought or oversold conditions in an asset price. |
| Relative Strength Index (RSI) | RSI is a technical analysis indicator that measures the magnitude of recent price changes to evaluate overbought or oversold conditions in an asset. It compares the average gains to average losses over a specified period and generates a value between 0 and 100. RSI is often used to identify potential trend reversals and to generate buy or sell signals. RSI > 70 typically indicates overbought conditions, while < 30 indicates oversold conditions. |  | Measures the strength of a stock's price action to identify overbought or oversold conditions. |
| Price-to-Earnings Ratio (P/E Ratio) | P/E ratio is a stock valuation ratio that compares a company's current share price with its earnings per share (EPS). It is calculated by dividing the market price per share by the EPS. The P/E ratio is used to assess a company's relative value compared to other companies in the same industry, and to determine whether a stock is overvalued or undervalued. |  | A measure of the market's expectation of a company's future earnings growth |
| Moving Average Convergence Divergence (MACD) | MACD is a trend-following momentum indicator that shows the relationship between two moving averages of a security's price. A nine-day EMA of the MACD called the "signal line," is plotted on top of the MACD line, which can act as a trigger for buy and sell signals. The MACD histogram represents the difference between MACD and its signal line. When the MACD is above its signal line, it indicates a bullish trend, and when it is below, it indicates a bearish trend. | MACD Line = 12-period EMA - 26-period EMA  Signal Line = 9-period EMA of MACD Line  Histogram = MACD Line - Signal Line | Identify trends and potential trend reversals in stock prices, using two moving averages and a histogram. |
| Ichimoku Cloud | Ichimoku Cloud is a technical analysis indicator used to identify potential trend reversals and gauge the strength of a trend. It consists of several lines that together form a "cloud," including the Conversion Line, Base Line, Leading Span A, and Leading Span B. The area between the Leading Span A and Leading Span B is shaded to represent support and resistance levels. The indicator is commonly used in conjunction with other technical analysis tools to identify buy and sell signals. | , plotted 26 periods ahead  , plotted 52 periods ahead  CL (Conversion Line) = Average of the highest high and lowest low over the past 9 periods  BL (Base Line) = Average of the highest high and lowest low over the past 26 periods. | Technical analysis of financial markets for trend identification and trading. |

**References**

[1] National Stock Exchange of India Limited (NSE), <https://en.wikipedia.org/wiki/National_Stock_Exchange_of_India>

[2] Why is there a difference between last traded price and closing price on NSE? By Zerodha, <https://support.zerodha.com/category/trading-and-markets/corporate-actions/general/articles/ltp-doesnt-match-close#:~:text=The%20last%20traded%20price%20(LTP,the%20actual%20last%20traded%20price>

[3] What Is NIFTY 50? How To Invest In NIFTY 50, <https://www.forbes.com/advisor/in/investing/what-is-nifty-50-how-to-invest-in-nifty-50/>

[4] Commodity Channel Index (CCI), Fidelity Investments: <https://www.fidelity.com/learning-center/trading-investing/technicalanalysis/technical-indicator-guide/cci>

[5] Relative Strength Index (RSI), Fidelity Investments: <https://www.fidelity.com/learning-center/trading-investing/technical-analysis/technicalindicator-guide/RSI>

[6] Price to Earnings Ratio (PE Ratio), Fidelity Investments: <https://www.fidelity.com/learning-center/trading-investing/technicalanalysis/technical-indicator-guide/pe-ratio>

[7] MACD, Fidelity Investments: <https://www.fidelity.com/learning-center/trading-investing/technical-analysis/technical-indicator-guide/macd>

[8] Investopedia: <https://www.investopedia.com/>

[9] The distribution of stock market returns, <https://klementoninvesting.substack.com/p/the-distribution-of-stock-market#:~:text=As%20you%20can%20see%2C%20on,the%20normal%20distribution%20relatively%20well>

[10] Index Funds vs Stocks, Key Differences: <https://smartasset.com/investing/index-funds-vs-stocks#:~:text=The%20biggest%20difference%20between%20investing,considerably%20greater%20chance%20of%20loss>

[11] Mallikarjuna, M., Rao, R.P. Evaluation of forecasting methods from selected stock market returns. Financ Innov 5, 40 (2019), <https://doi.org/10.1186/s40854-019-0157-x>