Analysis of investment options in the Indian market

by Manish Reddy

Submission date: 06-Dec-2021 08:58AM (UTC+0530)
Submission ID: 1721649594
File name: s_of_investment_options_in_the_Indian_market__Final_paper_1.pdf (294.68K)
Word count: 4158
Character count: 20970

Analysis of investment options in the Indian market

Manish Reddy M
Computer Science and Engineering
PES University
Bangalore, India
manishreddym22@gmail.com

Sethupathy R V
Computer Science and Engineering
PES University
Bangalore, India
sethupathyrv@gmail.com

Gaurav Jayaraj
Computer Science and Engineering
PES University
Bangalore, India
gauravssrvm@gmail.com

Likith B
Computer Science and Engineering
PES University
Bangalore, India
likithreddy20001@gmail.com

***Abstract*—Newcomers to the Investment market are burdened with a plethora of options to choose from. This paper aims to simplify this process by providing a comparative study and a comprehensive overview of different indices over a period of 20 years using data obtained form the national stock exchange. We weigh up Sensex and the Nifty indices and their numerous sub sectors and study the risk and return factors of various asset classes namely the fixed deposits, equities, real estate and bullion market. We also examine the internal and external factors affecting these classes. They are categorised into ones which are good for day trading and ones which make money in the long run. The correlation of these markets with the Indian economy is made using regression methods. We use time series analysis to model the growth of the classes and predict the ones which have the best returns on investment.**

***Index Terms*—Investment, Stocks, Economics, Time series analysis**

I. Introduction

India is the world's sixth largest economy by nominal GDP and third largest by power purchasing parity. Since 1991 India has recorded a historic GDP growth rate of 7% due to the increased economic liberalization marking the end of the License Raj. Most of this is mainly attributed to urbanization and the promotion of startup culture. We are moving away from agriculture sector and towards industrial and service sectors. The service sector constitutes up to 50% of GDP. In 2013, India surpassed China becoming the world's fastest growing economy up until 2018. India has a young population and a low dependency ratio. We are integrating evermore into the global economy. As such investors are bullish on the long term growth prospects of the markets. The government encourages participation of the populace in economic activities by providing various avenues to invest capital. As more people are brought out of poverty, the number of people willing to invest in the market rise as their disposable income increases.

Historically Indians invested mostly in Real Estate and gold in the form of jewellery. The stock market seemed too foreign and most of the populace is risk averse. But this is all set to change and most of this change can be seen presently. The GDP to stock market capitalization grew substantially over a period of 20 years. As the number of companies which issue IPO's increase the public will be spoilt for choice. This paper aims to address this issue.

Investment is necessary as money kept idle cannot beat inflation. As such there are 4 major asset classes which can be invested into.

- **Fixed Income Instruments:** Consists of bonds and fixed deposits issued by banks and private entities. There is minimal risk to investment and as such the returns are low. In fact the returns are so low that it hardly goes above the inflation. It is as good as no returns
- **Real estate:** Consists of buying and selling of commercial and non commercial land. This is long and complex process involving verification of documents and there is no official metric to quantify the gains made. It is also difficult to liquidize the investment.
- **Bullion market:** Constitutes trading of precious metals such as gold and silver. This is a safe method but the historical return isn't encouraging.
- **Equities:** Made up of trading shares of publicly listed companies on the Bombay Stock Exchange (BSE) and the National Stock Exchange (NSE). The main purpose of the stock market is to help facilitate their transactions. The participants can access any publicly listed company and trade from their point of view, as long as there are other participants who have an opposing point of view. This method generates the most CAGR(compound annual growth rate) but is also very risky. It is also difficult for new players to enter this market as it requires skill and patience.

Investing is becoming more accessible to the general public and is also encouraged. Even though most of the Indian economy is agrarian it is changing at a rapid pace. The urbanization of our country is happening rapidly and more sectors are incorporated in the market. People working in capitalist systems earn money working in corporations and the same is used to buy products made by these corporations. This is a self fulfilling cycle. Investing back in the same

organizations will keep this cycle well oiled.

II. Review of Literature

A. Macroeconomic Factors Do Influence Aggregate Stock Returns

The Aim of this paper [1] was to analyse the factors like macroeconomic developments exerting important effects on equity returns. Any variable that affects the future investment can be considered as a potential risk factor. The paper [1] shows that Macroeconomic variables are excellent candidates for these risk factors because they may simultaneously affect many firms' cash flows and may influence the risk-adjusted discount rate. Economic conditions during the time may also influence the number and types of investment opportunities available.

The study shows that higher inflation and money growth is usually looked on as a negative for stocks because it increases borrowing costs, increases input costs and reduces standards of living. But most importantly it reduces expectations of earnings growth, putting downward pressure on stock prices.

It focuses more on other variables that might have an impact of returns and risk evaluation like Consumer Price Index, Producer Price Index, Monetary Aggregate, Employment Report, Balance of Trade and Housing Starts. These were at the time ignored by most previous articles sought the time-invariant effects of macro innovations on equity returns.

The study shows evidence to support that Industrial Production growth is significantly positively correlated with real stock returns over the period 1926-1986 but not in the 1946-1985 sub period. This correlation could be due to economic news events which might explain the large stock market returns.

B. ARIMA Model- A statistical perspective for stock price forecasting

In [2] the author explains how researchers' desire to improve existing predictive models and build new ones to forecast stock prices because of the inclination of investors to cling to any forecasting methods that can guarantee huge returns and minimise risk.

Prevailing prediction models majorly fall under two perspective: Artificial intelligence and statistical techniques. Even though ANNs from the artificial intelligence perspective are renowned for their ability to learn complex patterns it was observed that ARIMA models from the statistical models vastly outperformed other techniques for short-term stock predictions.

The data set chosen by the the author is composed of four elements, namely: open price, low price, high price, close price. In the research, Closing price was chosen to represented as the price index to be predicted since it reflects all the activities of the index in the trading day.

To come up with the best ARIMA model, the author used the following criteria in the study after various experimentation:

- Relatively small BIC (Bayesian or Schwarz Information Criterion)
- Relatively small standard error of regression (S.E. of regression)
- Q-statistics and correlogram show that there is no significant pattern left in the auto correlation functions (ACFs) and partial auto correlation functions (PACFs) of the residuals, it means the residual of the selected model are white noise
- Relatively high value of adjusted R^2

The process of finding the right model for the given time series involves a general overview of whether the time series is stationary or not. If the ACF dies down extremely slowly, it means that the time series is non stationary and in that case it is converted into a stationary series by differencing. The model checking is done with Augmented Dickey Fuller unit root test. The value of d is chosen based on the level of differencing. The p and q value is chosen based on the above criteria after it is experimented upon several ARIMA models. If the model is good, the residuals of the model are a series of random errors and if there are no significant spikes of ACFs and PACFs ,it means that the residual of the selected ARIMA model are white noise and there will be no need to consider any AR(p) and MA(q) further.

The experimental results for Nokia and Zenith bank stock price prediction was discussed and the performance of the ARIMA model selected was quite impressive as there were some instances of closely related actual and predicted values.

C. Time Series Sales Forecasting using STL + ARIMA

The paper [3] uses a widely known approach to model time series data called Seasonal-Trend Decomposition using Loess(STL) and Autoregressive Integrated Moving Average(ARIMA). This model extracts the trend, seasonality and remainder components of the time series data and implements ARIMA to forecast the remainder component of the decomposed time series data. Then the seasonality component is added back in to complete the predictions.

STL method used here is an additive decomposition, as seasonal component(which repeats in certain known intervals) of the sales data does not vary greatly with trend(which are non-seasonal in the long term). The study found that there does not appear to be much trend in the data at all.

The paper [3] used Regression Decision Tree and Time-lagged Feed-Forward Neural Network to also forecast along with ARIMA but, the results weren't very different. The STL + ARIMA model seems to struggle with trends like moving holidays which greatly affect sales but was able to predict seasonal and remainder components with a decent accuracy.

In the decomposition of the data, Local Regression(Loess) smoothing technique is applied, determining how many distinct weeks(in a year) are combined to calculate the seasonal component. After this, the data seasonal and trend data are extended to forecast the test data. Then, to finish additive prediction model, ARIMA is applied to forecast remainder component. ARIMA takes into account the autoregression of a variable and its moving average. It also applies a technique

to transform the data into stationary data if necessary using the following equation.

$$ARIMA(p,d,q) = \sum_{i=1}^{p} y_i y_{t-i} + \varepsilon_t + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i}$$

In this equation, p is the number of terms included in the auto-regressive moving-window and q is the number of past residuals included in the moving-average.

III. Problem Statement and Proposed Solution

The stock market is said to move according to the emotions of the shareholders. There is no intrinsic value to a stock but it is what people assign to it. As such the stock market is highly volatile and uncertain. There are various factors that drive the price of the stocks. We develop the time series model using quantitative variables such as opening, closing, high, low and volume traded over a period of time. There are external factors which make the investors bullish or bearish on the markets. We model the data using quantitative variables and explain the anomalies using real world data. This paper also contains Machine Learning and time-series models that can forecast the stock prices based on historical data. A Comparative study of these models based on their accuracy and results with high accuracy making it evident that these values are indeed predictable to an extent.

A. Dataset

National Stock exchange(NSE) India publishes the nifty stock market index. The NIFTY50 is the coveted index consisting of the top 50 companies traded on the exchange. The other well known index is the SENSEX published by the Bombay Stock Exchange(BSE). These indices are a composition of many stocks from different sectors that collectively represent the state of the economy. They can also be used to benchmark a particular stock to see how it is performing against the market average. If a stock qualifies to be in a certain index it should meet the required criteria and if it fails it will get replaced by another stock which qualifies. Every stock is assigned a certain weightage in accordance with it's comparison with other stocks. The Indian market follows a free-float market capitalization in which the weights are assigned in proportion to the market capitalization of the company.

There are sub indices for different sectors. For example the consumer sector, the automotive sector and the banking sector. These indices are calculated as a subset of the NIFTY and SENSEX index based on the type of stock traded. We do a study on the growth of different sectors and predict the trend of future growth and decline.

The data is published for each day and the five parameters consisting of the opening price, closing prices, highest price, lowest price and the volume traded for each stock is listed on the index. The opening price represents the price of the stock at the start of the day and the closing price symbolises the price at the end of the day. The highest and lowest price signify their namesake for each day. The volume represents the amount of stock traded for each day. Another parameter called share turnover is also recorded which represents the amount of liquidity in stock. It is obtained by dividing the volume of stock traded by the number of shares outstanding and is directly proportional to liquidity of the company.

B. Data Preprocessing

The data set supplied by the National Stock Exchange contains various parameters with timelines ranging from the 2000s to the 2020s. It consists of 52 files in the form of Comma-Separated values (CSV). This includes 50 files containing historic data and trading volumes of 50 stocks of the NIFTY50 index. Aggregation of some specific companies was done along with their sector-wise indices. A common timeline of the data from the 2012 to the 2021 was chosen as this period is the intersection of all the stocks to be analysed. A wide portfolio of stocks consisting of companies in different sectors was used to provide a good balance where the boom of a particular industry doesn't impact the overall analysis significantly. The rows which contained null values were dropped. Some attributes were also dropped as they were more of an hindrance than of help. An estimation of the usefulness of the parameters was done to determine if it helps in the model prediction. A train test split of 80-20% was used for fitting and validating the data.

C. Data Fitting

In the following section, this paper will focus on all the forecasting techniques used to predict stock prices and the useful insights gained from those.

- **Linear Regression:** Linear Regression was the first model to be considered and surprisingly out performed most of the more complex models. The simple linear regression model predicted the test data with a maximum Root mean squared error(RMSE) of 46.6 with most of the RMSE values lying in the range of (1, 20).
- **Support Vector Regressor:** The next model to be considered was Support Vector Regressor. The model has a hyper parameter 'c' which is the regularisation parameter. The hyper parameter was optimised by simple linear brute forcing to minimise RMSE values. The model gives a decent RMSE compared to linear regression and even beats it for a few companies.
- **Decision tree Regressor:** The next model was Decision Tree Regressor. The model has a hyper parameter 'n' which stands for the maximum number of leaf nodes the tree can have. The parameter is again optimised by simple linear brute forcing to minimise RMSE value. Overall, this model performed very poorly compared to the Linear regression and SVR.
- **Random forest Regressor:** The next model is a Random Forest Regressor. The model is very similar to Decision Tree Regressor because it is an ensemble version of Decision Tree. They share the same hyper parameters which was fixed using a similar method. The model

overall performed better than Decision Tree Regressor and matched up to SVR for a few companies, but it never stood out as the best model for any of the companies.

- **SARIMAX:** A time series model called SARIMAX model was used in which the endogenic variable is set as Close price and exogenic variable is set as a combination of Open, Low, High, Volume and Turnover. The parameter tuning is done by grid-search. The model diagnostics such as ADF test(Augmented Dicky Fuller Test) is used to check if the time-series is indeed a stationary one and if detrending or differencing the dataset is necessary . Ljung Box Test was also used to see how good the model fits train data ,i.e ideally the residuals upto k-lags must follow white noise characteristic which our dataset did. The usefulness of the model was determined by using Theil's coefficient where we compare the MSE of Naive-Forecasting model with it. The model gives better RMSE compared to other models.

The volatility of stocks is measured using time tested historic measures α and β.

- α is the measurement of the deviation of the performance of a stock with respect to the market benchmark. It is given by

$$\alpha = R - R_f - \beta(R_m - R_f)$$

Where R is the returns for the company, R_f is the returns of a risk free investment. The value is taken as a rough estimate of the returns of an average Indian bank. R_m is the average value of returns of the entire stock market. Since this data is too vast to compute, nifty 50 was taken as a representative of the Indian stock market.

- β is the measurement of the variation of a particular stock with respect to the entire stock market. It is given by

$$\beta = covariance(R, R_m)/variance(R)$$

where R is the returns of a company and R_m is the returns of the entire stock market, once again represented by nifty 50.

An analysis tool called the exponential weighted moving average is used to guide newcomers about the state of the market. It smoothens out price data by creating a constantly updated average price. The average is taken over a specific period of time, like 10 days, 20 minutes, 30 weeks or any time period the trader chooses. We use this statistic to filter out the jagged price fluctuations.

IV. Results and Inferences

The stock market is affected by intangible factors and some of the historic ones that affected the SENSEX was demonetization, introduction of GST, 2014 election of the BJP government and the global recession of 2008. The stock market is also affected by periodic events such as budget declaration by the financial ministers and the end of the current financial year on March 31st. Monetary policy which is released by the Reserve Bank of India is a key event which determines the rate of inflation. The RBI meets every two months to determine this fiscal policy and the stock market wildly fluctuates during this time.

The statistical approach has proven to give very accurate predictions for the stock prices, in fact even better than traditional artificially intelligent models like ARIMA. The drawback is that these models alone in no way can understand the underlying complex dynamics of a stock market. These fail or give very poor results when we use a model that isn't particularly built for that time series. So every model has to be built for it's own particular use case. These models also fail to understand correlation between multiple companies since they are built only for a single company.

We use the values of α and β to provide a basic analysis of how good a stock is and serve as a factor to compare the stocks of multiple companies to find a stock most suitable for the users' needs.

- If the value of α is high and the value of β is also high, then the company is a good investment option for casual investors where there is a great opportunity for high returns coupled with a very high risk of losing the investment. Anyone interested in investing in such stocks need to follow momentum strategy to effectively gain higher returns.
- If the value of α is low and the value of β is high, then the company is a good investment option for day traders and people interested in short term gains.
- If the value of α is high and the value of β is low, then the company is an excellent investment option for anyone looking to earn passive income with little to no maintenance. This will be the ideal option for the majority of people looking to invest their money.
- If the value of α is low and the value of β is also low, then the company is not an investment option for anyone other than people looking to make money off dividends.

Using a scatter plot of the α and β for each company is a great way to spot outliers which might be good investments depending on the amount of time you might be willing to spend.

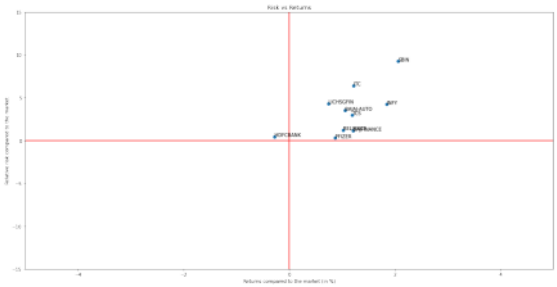

Fig. 1. Plot of Risk vs Returns for different stocks

There are two parts to analyze any particular stock - Technical and Fundamental analysis. A buyer needs to research on both the parameters to arrive at a conclusive evidence on whether to hold or sell. Fundamental Analysis is an ardours

and involved task. It consists of reading the annual reports of the company, P&L statement and the cash flows involved. Technical analysis is a much better option for beginners as it is pretty much the same across different stocks. This paper primarily used technical analysis to compare the stock prices. The stock market is said to repeat itself so using trusting alpha and beta to continue in the same fashion seems to be the best bet.

V. Conclusion

This project determined the various factors that influence the stock market focusing on the technical analysis. Most of the models implemented and found is done with respect to the New York Stock Exchange. This paper analysed the parameters affecting the NIFTY Exchange. Although future prediction of closing prices required the regression algorithm to contain the other parameters the Time series model SARIMAX gave the best prediction considering only the closing prices.

The investment market is a huge arena and entire fields of study are devoted to gamify the playing field. This paper provides a foray into it where we attempt to make predict the fluctuations. Sophisticated models like LSTM can be used to get better performances on the dataset. More research can always be done but we hope this paper makes a good starting point.

Acknowledgment

We thank our course instructors Dr Gowri Srinivasa and the courses' Teaching assistants for their guidance during each phase of the project. We also like to extend our gratitude to the Computer Science and Engineering Department at PES University for fostering a research culture and encouraging us to keep innovating.

References

[1] Flannery, M. J. (2002). Macroeconomic Factors Do Influence Aggregate Stock Returns. , 15(3), 751–782.edoi:10.1093/rfs/15.3.751

[2] A. A. Ariyo, A. O. Adewumi and C. K. Ayo, "Stock Price Prediction Using the ARIMA Model," 2014 UKSim-AMSS 16th International Conference on Computer Modelling and Simulation, 2014, pp. 106-112, doi: 10.1109/UKSim.2014.67.

[3] James J. Pao, Danielle S. Sullivan. Time Series Sales Forecasting. http://cs229.stanford.edu/proj2017/final-reports/5244336.pdf

[4] Ping-Feng Pai; Chih-Sheng Lin (2005). A hybrid ARIMA and support vector machines model in stock price forecasting. , 33(6), 497–505. doi:10.1016/j.omega.2004.07.024

[5] M, Hiransha; E.A., Gopalakrishnan; Menon, Vijay Krishna; K.P., Soman (2018). NSE Stock Market Prediction Using Deep-Learning Models. doi:10.1016/j.procs.2018.05.050

Contribution of team members

- Manish Reddy M (PES1UG19CS262):Compilation and consolidation of literature review and the final report. Domain level research and explored solution avenues. Analysed stocks and benchmarked it against the indices.
- Sethupathy R V (PES1UG19CS443):Data preprocessing and cleaning, exploratory data analysis, surveyed multiple research papers on this topic, Worked on building the models, Proof read the paper multiple times.
- Likith B (PES1UG19CS244):Data preprocessing and cleaning, exploratory data analysis. Did research into various models and methods for stock price prediction, Computation of α, β. Worked on dataset cleaning, Linear Regression model, SVR model and driver code for finding the best model for each company.
- Gaurav J (PES1UG19CS599): Data preprocessing and cleaning, exploratory data analysis. Hyperparameter tuning for Decision Tree Regressor, Random Forest Regressor and time series model and it's diagnostics

Appendix

Fig. 2. Comparison of AUTO sector index with 2 companies Bajaj and Tata stock data. It can be seen here that the index beat the above 2 stocks

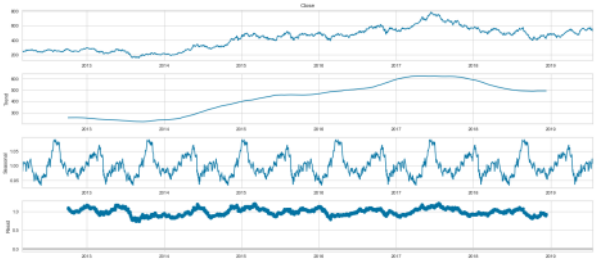

Fig. 3. Decomposition of the stock line into its respective components- Trend, Seasonality and Residuals.

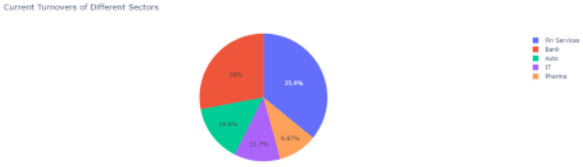

Fig. 4. Share of sectors in the market

Analysis of investment options in the Indian market

ORIGINALITY REPORT

19%	17%	10%	10%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	cs229.stanford.edu Internet Source	4%
2	www.researchgate.net Internet Source	2%
3	eprints.covenantuniversity.edu.ng Internet Source	1%
4	zerodha.com Internet Source	1%
5	Submitted to University of Durham Student Paper	1%
6	www.cnbc.com Internet Source	1%
7	www.coursehero.com Internet Source	1%
8	Submitted to PES University Student Paper	1%
9	academic.oup.com Internet Source	1%

10	Submitted to University of Witwatersrand Student Paper	1 %
11	www.sb.iub.edu.bd Internet Source	1 %
12	srustimanagementreview.ac.in Internet Source	1 %
13	www.fao.org Internet Source	<1 %
14	www.business-standard.com Internet Source	<1 %
15	Submitted to HELP UNIVERSITY Student Paper	<1 %
16	Submitted to University of Dundee Student Paper	<1 %
17	mafiadoc.com Internet Source	<1 %
18	www.ucstgi.edu.mm Internet Source	<1 %
19	"Data Science and Analytics", Springer Science and Business Media LLC, 2020 Publication	<1 %
20	Turan G. Bali, Robert F. Engle. "The intertemporal capital asset pricing model with	<1 %

dynamic conditional correlations", Journal of Monetary Economics, 2010
Publication

21	sciencepubco.com Internet Source	<1 %
22	www.financialexpress.com Internet Source	<1 %
23	Submitted to Tiffin University Student Paper	<1 %
24	Yanhui Song, Jiayi Cao. "An ARIMA-based study of bibliometric index prediction", Aslib Journal of Information Management, 2021 Publication	<1 %
25	Adebiyi A. Ariyo, Adewumi O. Adewumi, Charles K. Ayo. "Stock Price Prediction Using the ARIMA Model", 2014 UKSim-AMSS 16th International Conference on Computer Modelling and Simulation, 2014 Publication	<1 %
26	Flannery, M. J.. "Macroeconomic Factors Do Influence Aggregate Stock Returns", Review of Financial Studies, 2002. Publication	<1 %

Exclude quotes On Exclude matches < 5 words

Exclude bibliography On