INTERNATIONAL JOURNAL FOR INNOVATIVE RESEARCH IN MULTIDISCIPLINARY FIELD ISSN: 2455-0620 Volume - 6, Issue - 10, Oct – 2020
Monthly, Peer-Reviewed, Refereed, Indexed Journal with IC Value: 86.87 Impact Factor: 6.719
Received Date: 08/10/2020 Acceptance Date: 23/10/2020 Publication Date: 31/10/2020

Stock Price Trend Analysis Using Linear Regression

Aman Sharma
Student, MBA (Banking and Finance), SSBF, Pune, India
Email – amanjsharma8@gmail.com

Abstract: *In this paper we try to build up a strong model for picking stocks for short term trading, that is, from 1 day to 1 month. The purpose is predicting the prices and finding the estimation errors. We develop a basic mean model using linear regression which prompts a direct program and diminishes computational multifaceted nature of existing strong portfolio enhancement strategies. This paper tests the methodologies on genuine market information and talks about execution of the strong streamlining model exactly dependent on money related volatility or in terms of basic statistics Standard Deviation, and economic situation, for example, development, consistent state, and decrease in pattern.*

Key Words: *Regression, Shares, Mean, Volatility, Trend.*

1. INTRODUCTION:

If you analyse the economy it grows and declines over time, though not at a particular steady rate but it tends to fluctuate at various intervals as we have seen in the past and markets particularly the stock market shows that, in simple words if you have the economy growing, let's say the nominal economy, that is, the nominal GDP growing at 4% long-term corporate profits will also grow about at the same rate. In terms of stock prices, the basic difference between how the nominal economies are going and how the stock market is going, is the volatility that we see in stocks on a daily basis. That's the overall economy and that's the overall investment. Everybody thinks they are going to beat this there's that great gambling instinct in all of us that's why people watch financial news programs and trying to get information that might benefit their investment and grow their wealth. However, everybody cannot win at this game it'sgoingto average out but there is that hope that you might just beat the game. In order to beat the game, youhavetobeagainstthe consensus,itdoesn'tmeanthatyousimplycontradict everything. If you come up with an idea that is being taken into consideration by the consensus also it might be a good chance to make some money.

Securities exchange expectation and investigation is the demonstration of attempting to decide the future estimation of an organization stock or other money related instrument exchanged on a trade. Financial Exchange is the significant piece of economy of the nation and assumes an essential job in the development of the business and trade of the nation thatinthe endinfluences the economy of the nation. The two speculators and industry are engaged with financial exchange andneedstoknowwhethersomestockwillrise orfall over certain timeframe. The securities exchange is the essential hotspot for any organizationtoraiseassetsforbusinessextensions.Itdependsontheideaof interestandgracefully. Ontheoff chancethattheinterestforanorganization's stockishigher,atthatpointtheorganizationshare costincrementsandin the event that the interest for organization's stock is low, at that point the organization share value decrease.

In a large portion of the economies exchanging shares is a major business. From the data assembled from their site, it gives the idea that the stock intermediaries don't have any canny device which can help them exhortation their customers on which stocks are appropriate for them to purchase or sell. The predominant strategies show a pattern on future development of stocks and not the reasonable cost for any stock later on. It is in this manner desirable over have a device that doesn't simply point a heading towards cost development, yet in addition shows the most probable cost of the stock itself to a numerical worth that is anticipated by relapse. Where the objective qualities have been characterized as of now, relapse performs procedure on an informational collection. By including new data, the outcome can be broadened.

Is the market actually so arbitrary? Regardless of its clear multifaceted nature, the market, over the long haul, is driven basically by some predictable components. In spite of the fact that we can't see into the future, those variables offer a scope of market restores that are more unsurprising than you may might suspect.

The financial exchange isn't a fantasy machine. It can't generally give us what we need. Don't worryIf you'rea brand new capitalist and don't perceive these charts at first. After all, each roaring capitalist was a beginner at some point - there are ways to identify key shopping for opportunities on the charts, as well as essential signals that a stock ought to be oversubscribed.

To succeed you wish to find out sound and traditionally well-tried purchase rules and sell rules.

As you study these charts you will be able to see there are specific chart patterns that are repeated over and yet again whether or not in 1900 or 2000. This provides you with an enormous advantage once you learn to acknowledge these patterns show the impact that tell you once a stock is beneath skilled accumulation.

It is the distinctive combination of your finding stocks with huge will increase in sales, earnings and come on equity and sturdy chart patterns revealing institutional buying that along can materially improve your stock choice and temporal arrangement. The best professionals use charts and volatility to make money in intraday.

2. LITERATURE REVIEW:

A paper composed by Han Lok Siew and Md Jan Nordin exhibited the significance of utilizing ordinal or normalized values from a dataset made of heterogeneous sorts of information, in relapse forecast procedures. The paper speculated the elements to be considered for a forecast model, which conjugated the premise of performing basic examination of stocks. The factors were removed from accounting reports, pay articulations and other income proclamations given in the yearly reports of the specific stock tickers.

The information mining and its apparatus has assumed an indispensable job in investigating the information from various product houses. Utilizing information mining instruments and expository advancements we do a quantifiable measure of exploration to investigate new methodology for the venture choices .The market with immense volume of financial specialist with adequate information and have an expectation just as power over their speculations. The financial exchange some time neglects to pull in new speculator.

The explanation expresses that non-mindful and furthermore individuals would prefer not to approach to fall in to the hazard. A methodology with satisfactory aptitude is intended to assist financial specialists with ascertaining hidden examples from the notable information that have practical prescient capacity in their speculation choices. In this paper the NSE – Clever Midcap50 organizations among them top 4 organizations having max Midcap esteem has been chosen for investigation. The recorded information has a huge job in, helping the contributing individuals to get a diagram about the market conduct during the previous decade.

In the paper by varun arora in 2018 states that Securities exchange Expectation is the strategy for deciding future estimations of an organization's stock. Securities exchange Forecast has consistently pulled in individuals keen on putting resources into share market and supply of an organization for enormous benefits however it is hard to anticipate the stock estimations of an organization as it relies upon numerous components. Financial exchange continue differing step by step.

K.R. Sekar, K.S.Ravichandran and J.Sethuram composed a paper on the forecast of gold cost in the securities exchange dependent on a few free yet compelling factors. The expectation model utilized numerous direct relapse calculation to foresee the cost of gold in the market. Their model took a dataset comprising of verifiable gold costs, alongside different factors of numerous years on a month to month premise to take care of into their model which would be utilized for forecast later on.

Farhad, Seyyed Reza. In this paper, by applying straight relapse for foreseeing conduct of SP 500 file, we demonstrate that our proposed strategy has a comparative and great execution in contrast with genuine volumes and the investors can contribute privately dependent on that.

Strategic regression(LR), which is helpful for anticipating the event or non-event of a quality or result dependent on estimations of a lot of forecaster factors, is a multivariate investigation model [Lee, 2004]. In the zone of banking, corporate money and speculations, LR applications have as often as possible been utilized. For the default-expectation model, numerous specialists utilized Multivariate discriminant examination (MDA). Öğüt and Aktaş [2009] found that information mining procedures (ANN and SVM) are better appropriate to distinguish stock-cost control than multi variate factual strategies, for example, segregate examination or LR, since the exhibitions of information mining procedures as far as order exactness are better than those of multivariate procedures.

They proposed another double order strategy for anticipating corporate disappointment dependent on hereditary calculation, and proposed to approve its expectation power through observational examination. Minand Jeong [2009] thought about expectation precision with different techniques, for example, multi-discriminant investigation, calculated relapse, choice tree, and fake neural system and demonstrated that the parallel characterization strategy they proposed can fill in as a promising option in contrast to existing strategies for insolvency expectation.

3. MATERIALS:

Linear regression tries to model the connection between 2 variables by fitting a equation to determined information. One of the variable taken into account is instructive variable, and also the alternative is taken into account to be a variable.

At the core of a relapse model is the connection between two unique factors, called the reliant and free factors. For example, assume you need to figure deals for your organization and you've inferred that your organization's deals go here and there relying upon changes in Gross domestic product.

The business you are gauging would be the reliant variable on the grounds that their worth "depends" on the estimation of Gross domestic product and the Gross domestic product would be the free factor. You would then need to decide the quality of the connection between these two factors so as to conjecture deals.

In the event that Gross domestic product expands/diminishes by 1%, what amount will your business increment or lessening?

For instance, an analyst would possibly wish to relate the weights of people to their heights employing a simple regression model. Before making an attempt to suit a linear model to determined information, a modeler ought to initial verify whether or not or not there's a relationship between the variables of interest.

This doesn't essentially imply that one variable causes the opposite (for example, higher SAT scores don't cause higher school grades), however that there's some important association between the 2 A scatterplot is a useful tool in determinative the strength of the connection between 2 variables.

If there seems to be no association between the planned instructive and dependent variables (i.e., the scatterplot doesn't indicate any increasing or decreasing trends), then fitting a statistical regression model to the info most likely won't give a helpful model.

A valuable numerical live of association between 2 variables is that the coefficient of correlation, that may be a price between -1 and one indicating the strength of the association of the determined information for A statistical regression line has an equation of the shape Y = a + bX, wherever X is instructive variable and Y is the variable. The slope of line is b, and a is intercept (the price of y when x = 0).

The Mean Absolute Deviation (MAD) of a dataset is that the average distance between every data point and therefore the mean. It offers us an understanding regarding the variability during a dataset. Here's a way to calculate the mean absolute deviation.

Step 1: Calculate the mean.
Step 2: Calculate how far-off every data point is from the mean taking positive distances. These are referred to as absolute deviations.
Step 3: Add those deviations along.
Step 4: Divide the total by the quantity of data points.

The Mean Squared Error (MSE) tells us how close a regression curve is to a group of points. It is able to do this by taking the distances from the points to the regression curve (these distances are the “errors”) and squaring them. The squaring is important to get rid of any negative signs.

It additionally provides additional weight to larger variations. It’s referred to as the mean square error as you’re finding the common of a group of errors. The smaller the means square error, the nearer you're to finding the line of best match.

The Mean Absolute Percentage Error (MAPE) is an statistically used measure of understanding how correct a forecast system is. It measures this accuracy as a proportion, and also can be calculated as the average absolute % error for every period minus actual values divided by actual values. Where we see at is the actual worth and it is the forecast value, this is often given by:

$$\mathrm{M} = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right|$$

The mean absolute percentage error (MAPE) is the most widely used and common to forecast error, and it works best if there are no extreme values to the data (and no zeros).

4. METHOD AND ANALYSIS:

The question that we are trying to find the answer to is that using linear regression and other statistical tools can we predict the trend of stock prices or simply their movements?

Nature of the securities exchange is very unstable. The significant objective is to limit the vulnerability of the returns by carefully anticipating (as precisely as could be expected under the circumstances) the future stock costs and in this way distinguishing their changes ahead of time to relieve money related dangers. For the accompanying reasons we have created, our own expectation model.

The model uses historical volatility and Linear Regression using Excel as most of the analysis tools are hard to use and create, as they require a little understanding of coding. The model used by me is created in Excel itself and can be easily understood by anyone who has working knowledge of excel. The data being used is Secondary Data and is quantitative in nature. Sourced from BSE website and using data from January 2019-December 2020.

- Stock used is Tata Motors as it has been one of the most volatile stock during that period even though the Automobile sector was doing good
- Use scatter plot to get the Regression Line.
- The forecast is, Date multiplied by the slope line plus the intercept, which we get it using F(X) function in Scatter plot.
- We calculate Absolute Error using ABS formula in Excel which is =ABS (Forecasted value - The actual value), then calculating Squared Error which is just Squaring the Absolute Value. Mean Absolute Deviation and Mean Squared Error.

Then calculating percentage of error that is absolute error divided by the actual close. Mean Absolute Percent Error. As we can see in the image below how the model was created and looked in excel. It started with taking closing prices from the BSE website through their archive function which lets you access historical market price of every stock listed on the exchange. We have used historical market price for the year 2019-20 to get a complete understanding of the historical volatility for a year.

Once the prices are exported to excel, we have sort them as per the demands of the research. We use the last month as a sort of benchmark, the last months market price of the stock is compared with the findings in order to find out the Margin of Error. In our model we take historical volatility of each month in order to understand the latest volatility the stock is facing. We are basically trying to find a pattern using the historical volatility.

So, we have taken December month as the comparing factor once we predict the final results. We calculated the predicted prices using linear regression, we select the closing prices for the month of November and then plot a Scatter Diagram for the same that provides us with the F(x) equation which is calculated as per the movement or volatility that stock has shown for that month. How much did it move in a day in that month?

For example, the equation is (1.46x-17250) so here X is the next upcoming date. The model shows us how the price for that particular stock will move on the next upcoming date using the historical price movements for the previous month. That is the basics in our model in order to increase the accuracy of the future market price predicted we find the F(x) equation for every month using the Scatter plot in the order (Nov, Oct-Nov, Sep-Nov, Aug-Nov).

This helps in understanding how the stock has been moving for the previous months and helps in minimizing the Margin of Error. Once we have the readings for the previous four months, we take an Average of those four previous months and we get our Predicted Price. The Predicted Price is than compared to the Actual rates. In order to understand how viable is this data we do some further calculations to make our finding more accurate.

Further we calculate Mean Absolute Deviation, it helps us in understanding how much the price of the stock could move on a daily basis that helps us understand what could be the Low or High for that stock on a particular trading day. Then we also calculate Mean Squared Error or also known as Standard Deviation, we calculate this to see how Standard Deviation tells us about the volatility of the stock in short term, we understand that Standard Deviation is a good measure of volatility only when calculating volatility or long term.

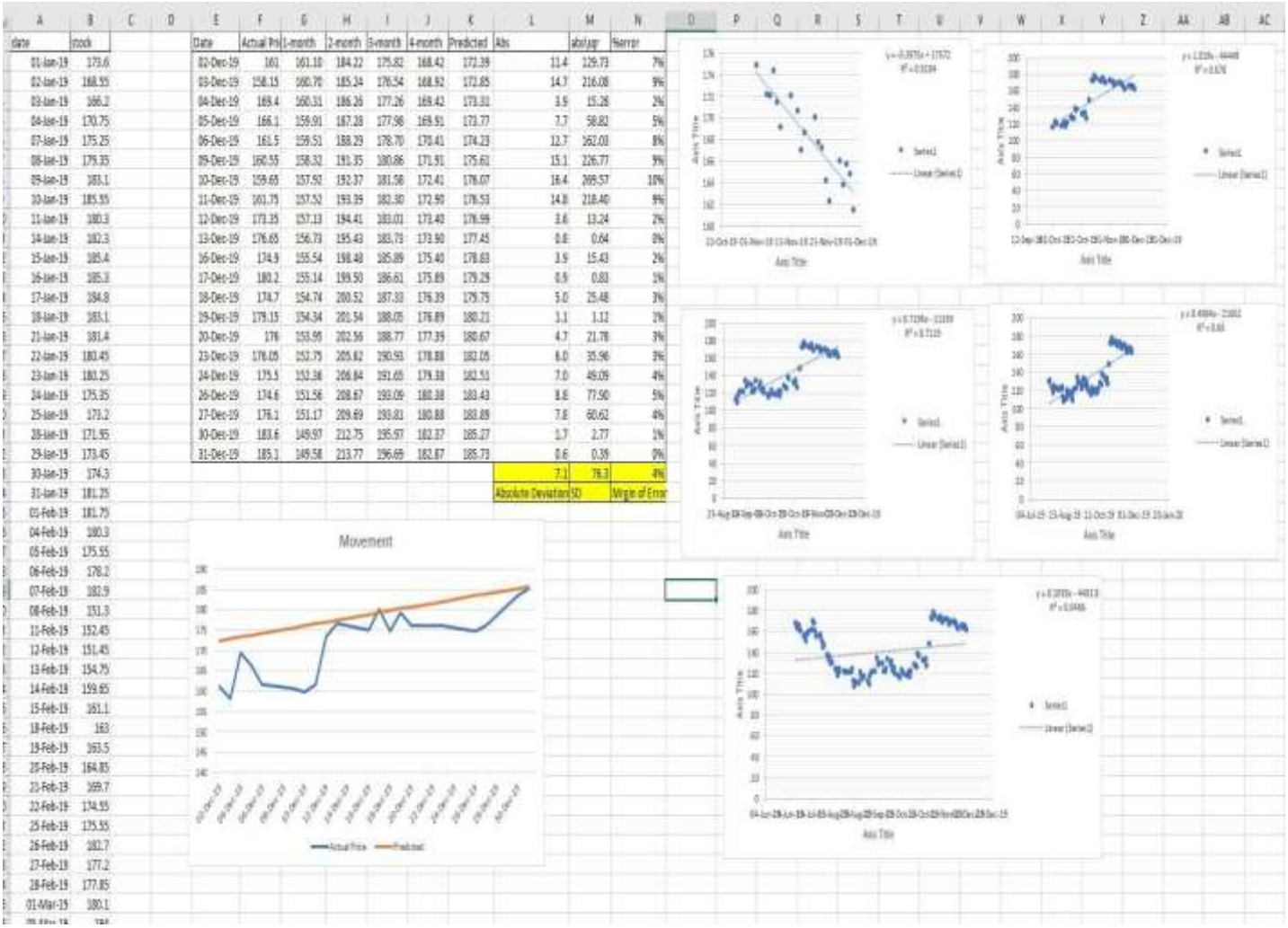

IMAGE-1 (model as on excel)

We also calculate Mean Absolute Percentage Error, it helps us in understanding our percentage of error, that is, how much is the percentage difference between the Predicted data and the Actual data. The data generated is not accurate but when used with other technical indicators can be very helpful to determine the possible trend of the stock in the short term.

So not intraday though but we could take small deals in the market say for a week or a month and make good profits using the information generated by the model.

The diagram below is a Scatter Plot Diagram it is a graph that helps us in understanding is there is any relationship between two variables. In our case the two variables are the Stock price and the date, and the relationship is derived by the equation Y as we can see in the graph below. We can also see that the graph provides us with the value of R Squared which helps us in understanding how powerful is the relationship between the two variables to get a better outcome when we are predicting the trend.

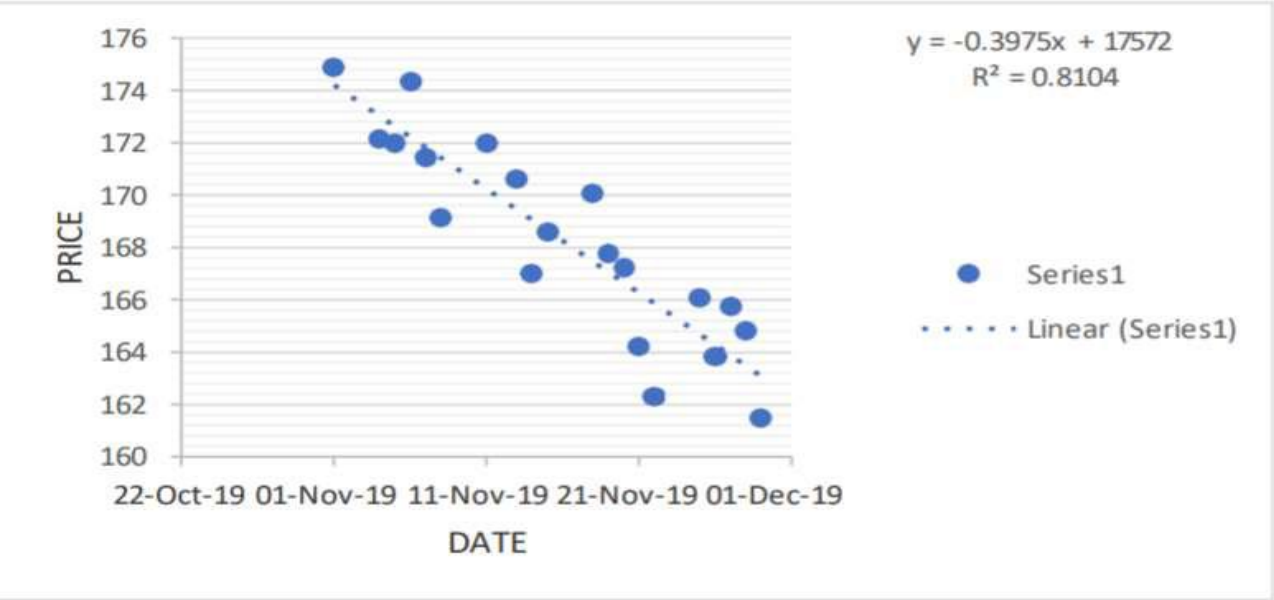

IMAGE-2(excel)

We get the MAD (Mean absolute Deviation) which is Rs. 7.1, that means MAD is a measure of absolute difference. How much the stock could move in a trading day or in a day or two as per historical volatility.

We get Standard Deviation but that is not much use here as SD is 76.3, SD's non-robust nature due to its usage of the sample mean in its calculation makes it unreliable. The absolute deviation provides a better picture for short term volatility and trend of a stock.

The Mean Absolute Percentage Error (MAPE) is a analytical measure of how precise can a forecast system be statistically. It measures this accuracy as a percentage, which is 4 % in our model for Tata Motors, which means that are predicted data is 96 % similar to the actual data and the Margin of Error is only 4%.

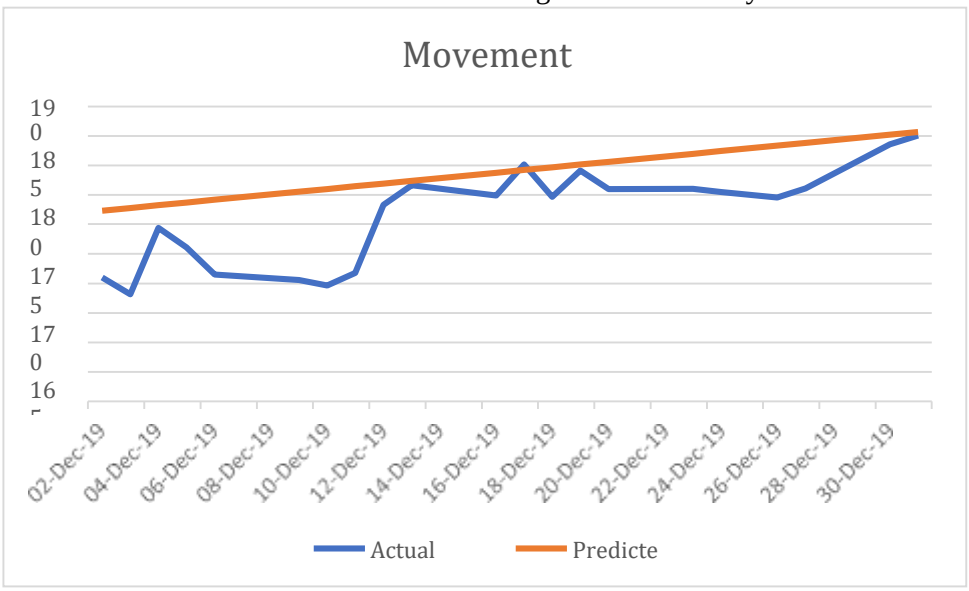

IMAGE-3 (excel)

As we can see in the above figure the Predicted Data is in a straight line and moving in one direction only, that is why, this model is currently a Trend predicter and not an actual stock price prediction model. But if you can predict the possible trend of a stock and also what is the error of percentage of your prediction you can take buying and selling decisions with possible Stop Loss as now you know how much possible profit and loss you can make by using the data predicted

5. FINDINGS AND LIMITATIONS:

The data collected can be used with other technical indicators to create a model that should be able to predict prices or possible trends of the stock prices. Linking other technical indicators with this model and create a more efficient model to predict the stock trend. To BUY or to SELL.

This model won't work in Turbulent times such as the 2008 Financial crisis or Freak events like CORONA Virus, because during such times Volatility is really High and the price fluctuations won't provide a clear picture. In our model we use Historical prices for months that are before for the month we have to calculate the prediction for. As we have seen in the current times how stock market has been falling from the month of march.

So, our model would show a very bizarre price pattern as there was no volatility the prices have just been falling down. So, the Volatility generated is the same as the predicted prices would show us a downward trend for the stock. However, with use of more technical indicators with this model we might be able to close the gap completely.
An unstable market environment will always raise problems for such models but with use of other indicators we can prepare for the worst also. Fundamental issues such as management related will not come in handy in this particular model as it is completely technical in nature.

6. CONCLUSION:

The data collected can be used with other technical indicators to create a model that should be able to predict prices or possible trends of the stock prices. Linking other technical indicators with this model and create a more efficient model to predict the stock trend. To BUY or to SELL.

This model can be used with the Monte Carlo Simulation too as that model also uses volatility to come up with possible future predictions using a very unique and random approach.

We can also use 1 min data of any stock and try this method to predict daily possible trend of a stock. This will help us in understanding the volatility in the stock prices on a minute basis.

Linking Technical analysis with Fundamental analysis and understanding the circumstances under which stocks moved the most. We might be able to find a pattern between the different type of analysis in order to understand the movement of the stock prices.

REFERENCES:

1. Investopedia, Economic Times, Bombay Stock Exchange, StatisticsToday, Khan Academy, Youtube.
2. Umadevi, B., et al. "[PDF] An Effective Time Series Analysis for Stock Trend Prediction Using ARIMA Model for Nifty Midcap-50: Semantic Scholar." *Undefined*, 1 Jan. 1970,
3. www.semanticscholar.org/paper/An-Effective-Time-Series-Analysis-for-Stock-Trend-Umadevi-Sundar/356879c2fc72465f5885315a16102975c6716226 (1)
4. Chaudhary, S., Arora, V., & Singh, V.K. (2018). REGRESSION BASEDSTOCK MARKET PREDICTION. *International Journal of Advance Research and Innovative Ideas in Education, 4*, 869-874. (2)
5. Siew, L. & Nordin, Md Jan. (2012). Regression techniques for the prediction of stock price trend. 1-5. 10.1109/ICSSBE.2012.6396535. (3)
6. K.R Sekar, "Gold price estimation using multivariable model", International Conference on Networks and Advances in Computational Technologies (NetACT) July 2017. (4)
7. Dr.P.K.Sahoo, Mr.KrishnaCharlapally, "StockPricePredictionUsing Regression Analysis". (5)
8. Sachin Kamley, Shailesh Jaloree R. S. Thakur, "A Linear Regression Approach To Prediction Of Stock Market Trading Volume:A Case Study" (6)
9. Lee, S. 2004. Application of likelihood ratio and logistic regression models to landslide susceptibility mapping using GIS. Environmental Management 34(2), 223-232. (7)
10. Öğüt, Hulisi, et al. 2009, Detecting stock-price manipulation in an emerging market: The case of Turkey, Expert Systems with Applications 36(8),
11. 11944-11949.
12. Min, Jae H., and Chulwoo Jeong. 2009. A binary classification method for bankruptcy prediction, Expert Systems with Applications 36(3),5256-5263. (9)
13. Gharehchopogh, F. S., & Khalifehlou, Z. A. (2012). A New Approach in Software Cost Estimation Using Regression Based Classifier. AWER Procedia Information Technology and Computer Science, Vol: 2, pp. 252-256. (10)
14. Draper, N. R., Smith, H., & Pownell, E. (1966). Applied regression analysis (Vol. 3). New York: Wiley (11)