STOCK MARKET FORECASTING USING LASSO, LINEAR REGRESSION, RIDGE REGRESSION AND GARCH NEURAL NETWORK

Ademola Abiodun Saheed
Baum Tenper Research Institute
saheed.ademola313@gmail.com

Akinyemi Mary
Department of Statistics
University of Lagos
maryi.akinyemi@gmail.com

Chika Yinka-Banjo
Department of Computer Science
University of Lagos
chikagog@gmail.com

ABSTRACT

The capacity to properly estimate stock reward patterns, risks, and returns is the most disputed subject in the financial business. Every stock investor strives to maximize profits while limiting or avoiding risk. Machine learning has sparked a surge in research interest. Several studies have been conducted to investigate the predictability of returns and incentives using machine learning methodologies. While the majority of previous work was focused on benchmarking the overall performance of particular research algorithms. Along with Linear regression, LASSO regression, Ridge regression, Elastic Net and GARCH Neural Network, we investigate a much broader range of styles in our work. In addition, unlike most other research, we investigated the difference in accuracy among all of the algorithms utilized in our study. Based on the Linear, Ridge, LASSO, and Elastic Net evaluation criteria. LASSO was discovered to outperform the other methods. Various GARCH varieties were tested, and AVGARCH(1,1) was shown to have the lowest Akaike information criterion (AIC) and Bayesian information criterion (BIC). AVGARCH(1,1) was then combined with the LSTM to enhance price volatility prediction.

***K*eywords** Linear regression, LASSO, Ridge regression, Elastic Net, GARCH, AVGARCH, LSTM

1 Introduction

Every stock trader, investor, and bank seeks to maximize returns while avoiding risk. Researchers have researched various new strategies over time in order to develop solid, viable alternatives to outperform the market and maximize earnings. With the introduction of machine learning and big data, they have developed an exceptional capacity to manage large amounts of data, prompting academics to seek new machine-learning methodologies in asset management and the stock market.

The goal of this study is to look at and forecast stock market dynamics utilizing basic and sophisticated machine learning and econometric methodologies..

This paper is divided into five (5) sections. First, the introduction and problem statement of the research are presented; the second section reviews some literature; and the third section is the methodology of the research, which includes a thorough examination of Ordinary Least Squares (OLS), Ridge, LASSO, Elastic net regression techniques, and Generalized Autoregressive Conditional Heteroscedasticity (GARCH). For regression reasons, their most prevalent strengths and weaknesses will be discussed. The fourth section gives an empirical research in which selected models are tested using real stock historical data, each models are evaluated, and the estimation results are explained. The last section concluded the research

1.1 Problem Statements

One of the most contentious issues in the business and financial sector is the capacity to forecast the overall direction or trend of stock rewards, risks, and returns. While the majority of financial analysts, economists, and statisticians have always been charged with building and improving existing models for forecasting stock market returns, the outcomes of their collective efforts have been equivocal. The Efficient Market Hypothesis underscores the premise of the stock market's incapacity to precisely predict risks and rewards, as well as returns (EMH). According to this concept, When fresh knowledge is disclosed, it spreads like wildfire, and big capital markets are swift and actively efficient in absorbing and incorporating this new knowledge and ideas into their markets. This results in an unanticipated price change since it is difficult to foresee the price change at time $t+1$ based on current knowledge (i.e. at time t). The price at $t-1$ already reflects the existing price at the moment. This means that changes in price at time $t+1$ will only be reflected at time $t+1$ as a result of news released at $t+1$.
The efficient market hypothesis was first used by [1]. He recognized that stock price fluctuations are equivalent to a random walk. According to the random walk model utilized in this notion, subsequent price changes are similar and independently distributed random variables. According to [1], these properties of stock prices can be compared to efficient markets. H. A competitive marketplace where pull and push forces between a large number of rational participants make it such that real pricing always completely reflect all available information. Using the classification criteria proposed by Roberts (1967),[2] contrasted his three forms of market efficiency:

•The susceptible shape efficiency: In this marketplace setting, the quantity and pleasant of to be had statistics set well-known shows ancient fees only. With this, the destiny fees of shares can't be decided with the aid of using reading fees from the past;
• The semi-sturdy form: This arise while fees handiest conforms to all publicly to be had information (e.g. bulletins of inventory splits, annual earnings, etc.);
• The robust form: that is a state of affairs in which modern inventory expenses mirror all statistics regardless of the forms; each personal and public (e.g. statistics made to be had handiest to company executives).

In [3], the set of data from the vulnerable for performance became later elevated to account for and encompass different unbiased variables which include dividend yields and hobby rates. As a result, a vulnerable shape of performance remained a fashionable region of check for returns and rewards predictability. Studies have found out that the green marketplace speculation isn't always universally accepted. Many academics have discovered that there are numerous market inefficiencies (i.e. anomalies) that might also additionally arise periodically, which suggests that there may be a slight probability that inventory charges are predictable ([4]; [5]). Predictions of inventory charges are typically mentioned withinside the literature: essential evaluation and chartist or technical theories. Fundamental evaluation is centred on measuring the inherent cost of an inventory with the aid of using investigating the simple determinants that effect the income electricity of the company in phrases of monetary and business conditions, control quality, etc. The predominant goal of the use of this technique is to decide the location of the real fee of an inventory as to be greater or much less than its intrinsic cost. For instance, whilst the real fee methods of the intrinsic cost, figuring out the intrinsic cost of an inventory is equal to figuring out the destiny expenses of the equal inventory ([6]). Technical evaluation, on the alternative hand, is taken into consideration an appreciably one of a kind method for predicting inventory returns. This method proposes that repeating fluctuations may be regarded from preceding historic marketplace information (maximum especially, records concerning fee and information volume). More formally, the technical evaluation assumes that successive fee adjustments are depending on previous ones [[7]]. This means that via way of means of figuring out successive fee adjustments, records from preceding expenses is needed. As a result, at each point in time, the collecting of historical rate changes is critical in projecting future rate changes. [8] additionally once helps with the usefulness of the technical evaluation. However, a long time ago, maximum prediction fashions that appoint the use beyond and ancient statistics on fee information and different variables as inputs had been primarily based totally on traditional statistical strategies such as linear and regression patterns, autoregressive (integrated) transferring average, and so on. But, inventory charges are not often linear and include random noise, non-stationary (imply and variances aren't steady for any decided on pattern length throughout the information set), and show off non-linear features which can be past the scope and can't be defined with the aid of using easy linear fashions ([9]). To cope with this anomalies, numerous device mastering strategies had been evolved to enhance prediction results.
In examining these, the technical evaluation technique could be used primarily, totally based on the belief that preceding and present inventory facts contain, beneficial data. This examine could be done using 3 trendy studying algorithms: LASSO (least absolute shrinkage and choice operator), Ridge Regression, ElasticNet and GARCH Neural Network. Comparisons could be made approximately at the end result from those algorithms and legitimate inferences and conclusions primarily based totally at the effects of the evaluation could be drawn for generalization purposes.

2 Literature Review

Machine learning (ML) is the field of research that allows computers to learn automatically from experience without the need of commands [10]. Recently, ML-primarily based totally strategies have attracted constantly growing studies interests. This is because to their ability to deliver modern results in a variety of areas such as language comprehension, vision, and speech recognition. In the commercial enterprise and economic sector, regardless of the overall notion withinside the green marketplace hypothesis (despite the fact that now no longer believed through all), numerous varieties of studies had been carried out to research the predictability of inventory marketplace returns and rewards the usage of a few advanced, supervised and innovative ML strategies. For instance, [11] carried out a evaluation check at the overall performance of 3 ML algorithms. These had been utilized in figuring out whether or not there's a opportunity that a one-day-beforehand go back of all SP 500 index components will outperform the marketplace or now no longer. A pattern of 31 lagged easy returns turned into randomly decided on as enter variables (i.e. explanatory variables) and earnings generated through a buying and selling approach had been used as an assessment metric. They found out that deep neural networks and gradient-boosted timber had been outperformed through random forests. They additionally found out that an similarly weighted mixture of these 3 algorithms yielded higher outcomes than while used independently. [12] done a look at geared toward distinguishing among the overall performance and abilties in addition to weaknesses of econometric models (GARCH, AR, and ARIMA) with Machine Learning methods (synthetic neural networks and SVM) in forecasting and figuring out positive/poor returns of 34 monetary indices that consist of each rising and evolved markets. Using easy covariates primarily based totally on open, high, low, and near prices, with a few technical signs such as (easy) shifting averages, relative electricity index, shifting common convergence-divergence, Williams $\%R$ and accumulation distribution oscillator, they determined that the general overall performance of the pleasant ML algorithm (SVM) outweighs the pleasant econometric method (AR) for each one-hour-beforehand and one-day beforehand forecasting windows. However, they also stated that technical indicators give little or no advantage over simple price-based totally variables.

While maximum of the researchers centered on forecasting short-time period windows, particularly day by day and weekly predictions, [13] centered on figuring out whether or not the one-year-in advance inventory fee of 5767 European groups will upward push at a predetermined threshold (15%, 25%, and 35%). eighty monetary and macroeconomic signs had been integrated as enter variables and primarily based totally at the region protected under the receiver running feature curve (AUC), they arrived at a end that random woodland is the pleasant acting set of rules preceded via way of means of SVM, kernel factory, adaptive boosting, neural networks, k-nearest neighbours, and ultimately LASSO regression. However, while the bulk of preceding works got down to benchmark the overall performance of only some gadget studying algorithms, in our have a look at we study a much wider set of fashions along with ridge regression, LASSO regression and Artificial Neural Networks. Moreover, in nearly all preceding studies, the authors did now no longer behavior suitable statistical checking out which will verify if the suggested consequences are statistically considerable or now no longer. On the contrary, we carry out forms of speculation checking out which will determine the statistical importance of our findings. First, we examine if the anticipated accuracy of each strategy is statistically greater than random guessing. Second, we investigate whether or not the difference in accuracy discovered between the two exceptional techniques is statistically significant. Finally, in evaluation to the prevailing literature in which every have a look at commonly covers most effective one prediction window, We examine how the overall performance of our styles differs as a result of three forecasting horizons: one-day ahead, one-week ahead, and one-month ahead.

3 Methodology

3.1 Data description

The statistics used for the studies is the rate records and buying and selling volumes of the fifty inventory withinside the index from NSE (National Stock Exchange) India and NSE30. All datasets are at an afternoon degree with pricing and buying and selling values. The statistics spans from 1st January, 2000 to thirtieth April, 2021.

3.2 Data Pre-processing

The data price characteristics are translated into daily stock returns, and the time series is divided into a training and test set. Converting stock prices into daily stock returns improves the dataset's stationarity.

- Performing log transformation
- Obtaining the Volatility of the prices

We created a time series plot of the initial price to show the price's stationarity as well as its dispersion. After the pre-processing, the data was re-visualized to see the changes.

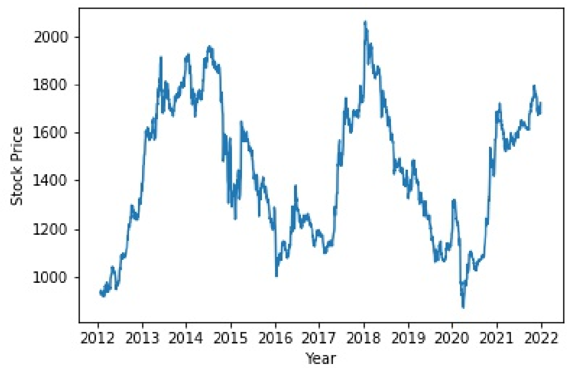

Figure 1: Time series plot of the original data

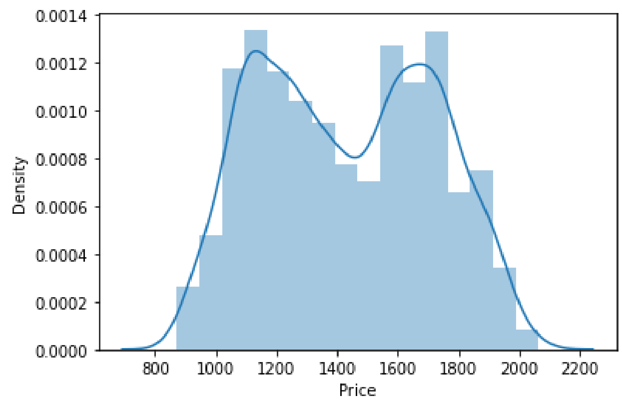

Figure 2: Distribution plot of the original data

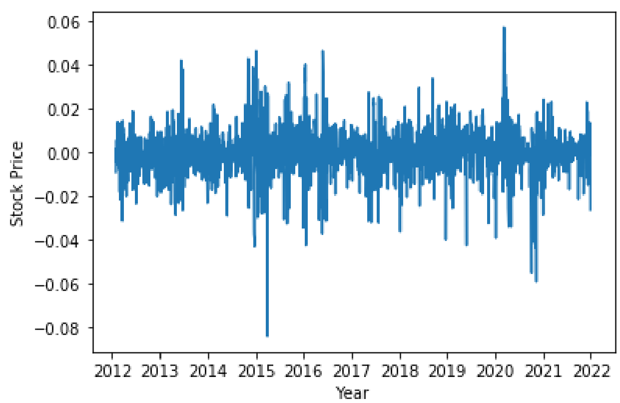

Figure 3: Time series plot of the log transformed data

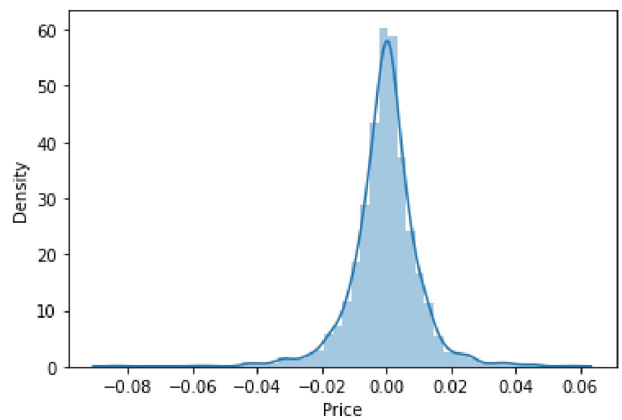

Figure 4: Distribution plot of the log transformed data

3.3 Linear Regression

Each $x_i = (x_1, \ldots, x_i)$ is a p-dimensional vector of features, and each y_i is the related response variable in Linear. The goal is to use a linear combination of characteristics to estimate the response variable y_i.

$$y_i = \beta_0 + \sum_{j=1}^{p} x_{ij}\beta_j \tag{1}$$

The vector of regression parameters $(\beta_1, \ldots, \beta_p)\varepsilon\mathbf{R}^p$ and an intercept term of $beta_0$ are used to parameterize the linear regression model. Ordinary least square regression is defined as linear regression with a loss, $\ell(u, v) = (u - v)^2$. The cost function must be optimized in order to assess the best fit line.

$$argmin_{\beta_0,\beta}\ell(\beta) = \frac{1}{N}\sum_{i=1}^{N}\left(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j\right)^2 \tag{2}$$

Which is equivalent to

$$argmin_{\beta_0,\beta}\frac{1}{2}||X\beta - y||_2^2$$

The optimal solution, β of 2 is given by

$$\beta = (X^T X)^{-1} X^T y \tag{3}$$

3.4 Ridge Regression

Ridge regression adds a factor to the cost function 2 called "sum of squares of the coefficients." Ridge regression employs a penalty or regularizer of the kind $||.||_2^2$ (i.e. $l_2 norm$). It essentially does it to try to minimize the sum of the error term together with the sum of squares of coefficients which we aim to find.
As a result, the optimization problem may be written as

$$\min_{\beta_0,\beta} \ell(\beta) + \lambda||\beta||_2^2$$

. i.e

$$argmin_{\beta_0,\beta} \left\{ \frac{1}{N} \sum_{i=1}^{N} \left(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j \right)^2 \right\} + \lambda \sum_{j=1}^{p} \beta_j^2 \tag{4}$$

Where λ is the regularization coefficient and $\lambda \geq 0$ determines the degree of shrinkage, i.e. the bigger λ, the more shrinkage and the coefficients are more resilient to collinearity + model prone to overfit.
The following closed form solution

$$\beta = (X^T X + \lambda I)^{-1} X^T y$$

determines the coefficient values of the aforesaid cost function.

3.5 LASSO Regression

A regularization term of "sum of the absolute value of the coefficients" is introduced to the cost function 2 in LASSO regression. LASSO employs a penalty or regularizer called $||.||_1$ (i.e. $\ell_1 - norm$). LASSO directly conducts feature selection by trimming the coefficients of superfluous variables to zero. Its optimization problem's mathematical formulation

$$\min_{\beta_0,\beta} \ell(\beta) + \lambda||\beta||_1 \tag{5}$$

As a result, the coefficients (parameters) are penalized for being sparse. The quantity of sparsity is controlled by the hyper-parameter $\lambda \geq 0$. The bigger the λ, the greater the sparsity, and the less likely the model is to overfit. Thus, in generalized form, the cost function for both ridge and LASSO is given by:

$$E(\beta) + \lambda R(\beta) \tag{6}$$

The error term is $E(\beta)$, and the regularization term is $R(\beta)$. LASSO directly conducts feature selection by reducing the coefficients of redundant variables to zero. Ridge, on the other hand, decreases the coefficients to arbitrarily low but not zero values. Because LASSO regression produces a sparse solution, many of the model coefficients become absolutely zero, $\beta = 0$. If $\beta*$ is the best model obtained, which is given as

$$\beta^* = argmin[E(\beta) + \lambda R(\beta)] \tag{7}$$

The $argmin$ function finds β values for which the equation 6 is the smallest.

3.6 Elastic Net

As a regularizer, Elastic Net employs a convex combination of ℓ_1 and ℓ_2 $norms$, with a parameter $0 \leq r \leq 1$ known as the $l1_{ratio}$.
The mathematical formulation of the optimization problem is given as

$$\min_{\beta_0,\beta} \ell(\beta) + \lambda r||\beta||_1 + \lambda\frac{1-r}{2}||\beta||_2^2 \tag{8}$$

This formulation penalize both size and sparsity of the coefficients (parameters) and tries to get the best of both worlds of ridge regression and LASSO regression. Elastic net is useful where multiple features are correlated with one another. $r\varepsilon[0, 1]$, controls trade-off between Ridge regression and LASSO. When $r = 1$, the optimization problem yields a ridge regression; while $r = 0$ yields LASSO.

3.7 Generalized Autoregressive Conditional Heteroscedaticity (GARCH)

Consider a stock market index I_t and its rate of return r_t, which we can calculate as $r_t = \frac{I_t - I_{t-1}}{I_{t-1}}$[14]. The index t represents daily closing observations. Instead of return, rate of return was employed to reduce the data dimension. The conditional distribution of the sequence of disturbances that follows the GARCH process may be expressed as

$$\varepsilon_t | \psi_{t-1} \sim N(0, h_t)$$

where ψ_{t-1} signifies all accessible information at time $t-1$. The conditional variance ht is

$$h_t = w + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j h_{t-j} \tag{9}$$

where $p \geq 0$, $q > 0$ and $w > 0$, $\alpha_i \geq 0$, $\beta_j \geq 0$ for non-negative GARCH (NG(p,q)) process. When $p = 0$ and at least one of the ARCH parameters is non-zero $(q > 0)$, the GARCH(p,q) model simplifies to the ARCH(q) process, [14]. The GARCH regression model for the r_t series is expressed as

$$\Phi_s(B) r_t = \mu + \varepsilon_t \tag{10}$$

with $\Phi_s(B) = 1 - \Phi_1 B - \ldots - \Phi_s B^s$

$$\varepsilon_t = \sqrt{h_t} e_t$$

$$e_t \sim N(0,1)$$

$$h_t = w + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-1}^2 + \sum_{j=1}^{p} \beta_j h_{t-j} \tag{11}$$

where B denotes the backward shift operator $B^k y_t = y_{t-k}$. The parameter μ represents a constant term, which is often considered to be near to or equal to zero in practice. The order of s is frequently 0 or tiny, suggesting that there are no chances of forecasting r_t from its own past. In other words, r_t never has an auto-regressive process. The w, α_i, and β_j parameters can be left unconstrained, giving the unconstrained GARCH (UG(p,q)) model. If the parameters are limited in such a way that

$$\sum_{i=1}^{q} \alpha_i + \sum_{j=1}^{p} \beta_j < 1$$

Because the mean, variance, and autocovariance are limited and constant across time, they imply the weakly stationary GARCH (SG(p,q)) process. When the model is integrated in variance, the multistep predictions of variance do not always approach the unconditional variance; that is,

$$\sum_{i=1}^{q} \alpha_i + \sum_{j=1}^{p} \beta_j = 1 \tag{12}$$

There is no unconditional variance for the IGARCH model. It is worth noting, however, that the integrated GARCH or IGARCH (IG(p,q)) model can be strongly stationary even if it is not weakly stationary (Nelson, 1990). Nelson proposed the exponential GARCH or EGARCH (EG(p,q)) model (1991). According to Nelson and Cao (1992), the non-negativity restrictions in the linear GARCH model are overly severe. The GARCH model puts non-negative constraints on the parameters α_i and β_j, whereas the EGARCH model has no restrictions on these values. The conditional variance, h_t, in the EGARCH model is an asymmetric function of lagged disturbances, ε_{t-1} :

$$\ln(h_t) = w + \sum_{i=1}^{q} \alpha_i g(Z_{t-1}) + \sum_{j=1}^{p} \beta_j \ln(h_{t-j}) \tag{13}$$

where

$$g(Z_t) = \theta Z_t + \gamma[|Z_t| - E|Z_t|]$$

$$Z_t = \varepsilon_t | \sqrt{h_t}$$

The coefficient of the second term in $g(Z_t)$ is set to be 1 $(\gamma = 1)$ in this formulation. Note that

$$E|Z_t| = \left(\frac{2}{\pi}\right)^{\frac{1}{2}}$$

if Z_t $N(0,1)$.The GARCH-M ((G(p,q)-M) model has the added regressor that is the conditional standard deviation.

$$r_t = \mu + \delta\sqrt{h_t} + \varepsilon_t \tag{14}$$

$$\varepsilon_t = \sqrt{h_t}e_t$$

where h_t is subject to the GARCH process [14]. The GARCH models are frequently used for analyzing stock market indexes. Furthermore, because a short lag in the GARCH model is adequate to describe long-memory processes with shifting variance (French et al., 1987; Franses and Van Dijk, 1996), the performance of GARCH models is assessed in this work by employing SG(1,1), UG(1,1), NG(1,1), G(1,1)-M, EG(1,1), and IG(1,1).

3.8 Experimental set-up and result

There are many assessment metrics that exist for comparing prediction on regression Machine Learning algorithms. To examine our models, Root Mean Square Error(RMSE), R^2 and Mean Absolute Error (MAE) were used.

Table 1: Description of the original dataset

Index	Time Period	N	Open	High	Low	Mean	SD
NSE 30	2012.01 -2021.12	2457	1432	1440	1424	1433	285
NSE India	2012.01 -2021.12	2456	32143	32317	31990	32171	6289

Table 2: Descriptive of Log returns data

	N	Mean	sd	Min	Max	Skewness	Kurtosis	S.E
NSE30	2456	-0.02	1.03	-8.08	5.86	-0.22	5.29	0.02
NSE India	2455	-0.02	0.99	-7.67	5.16	-0.23	5.37	0.02

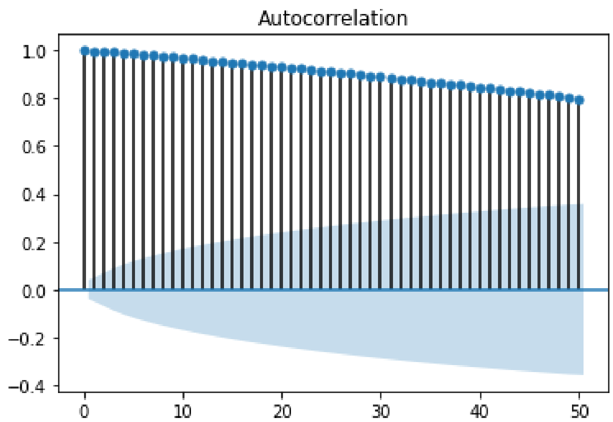

Figure 5: The ACF visualization of the original data with a lags value of 50

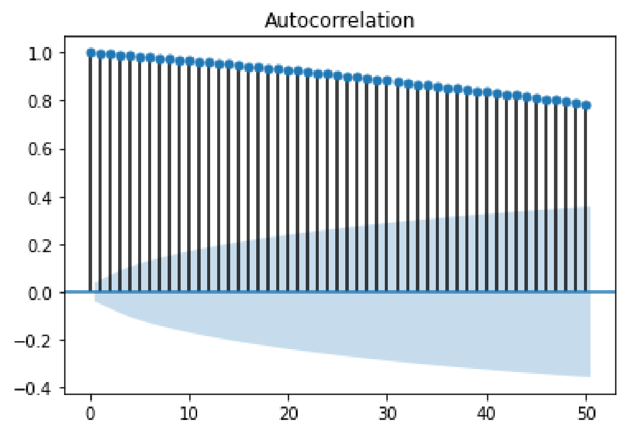

Figure 6: Time series plot of the transformed data

The shaded place is the self assurance with a default value of $\alpha = 0.05$ Anything inside this variety represents a fee that has no considerable correlation with the maximum current fee for the price. From our inventory records for NSE 30, we ought to examine that preceding costs impact the cutting-edge price.

The strength of this relationship is measured on a scale of -1 to 1 when -1 being a 100

Table 3: Comparison of the results of Linear, LASSO, Ridge, Elastic-Net and GARCH models on the train set.

Data	Model	RMSE Train	R^2 Train	MAE Train
NSE 30	Linear Regression	112.07	0.85	87.82
	LASSO	112.09	0.85	87.80
	Ridge	112.07	0.85	87.82
	Elastic Net	112.09	0.85	87.80
NSE India	Linear Regression	2405.70	0.85	1876.45
	LASSO	2405.90	0.85	1876.10
	Ridge	2405.70	0.85	1876.45
	Elastic Net	2405.89	0.85	1876.24

Table 4: Comparison of the results of Linear, LASSO, Ridge, Elastic-Net and GARCH models on the test set.

Data	Model	RMSE Test	R^2 Test	MAE Test
NSE 30	Linear Regression	116.70	0.83	91.40
	LASSO	116.44	0.83	91.25
	Ridge	116.70	0.83	91.40
	Elastic Net	116.43	0.83	91.25
NSE India	Linear Regression	2473.78	0.84	1888.75
	LASSO	2374.45	0.85	1860.38
	Ridge	2473.88	0.84	1888.76
	Elastic Net	2375.63	0.85	1862.27

From the assessment of the consequences of Linear, Ridge, LASSO and Elastic Net, with extraordinary assessment criteria, LASSO is the excellent performer for the extraordinary inventory records used.

Table 5: Box-Ljung test

		NSE 30	NSE India
Box-Ljung test (for returns)	test value	19.860	18.129
	p-value	0.0013	0.0027

Box-Ljung check desk allows us to test whether or not the rate of returns has an ARCH influence. The null hypothesis states that the rate of return has no effect on ARCH, whereas the alternative hypothesis states the opposite.
Based on a 5% significance level, all p-values are less than 0.05, demonstrating that return pricing has an ARCH affect.

Table 6: The results of all volatility models for NSE India

	GARCH (1,1)		AVGARCH (1,1)		GJR-GARCH (1,1)	
	Normal	Student-t	Normal	Student-t	Normal	Student-t
Omega	0.085	0.1238	0.0885	0.1121	0.0844	0.1261
S.E	3.22e-02	3.593e-02	3.345e-02	3.208e-02	3.187e-02	3.783e-02
p-value	7.991e-03	5.706e-04	6.405e-03	4.719e-04	8.076e-03	8.601e-04
Alpha	0.2197	0.3446	0.2127	0.2922	0.2343	0.3331
S.E	5.085e-02	6.283e-02	3.980e-02	4.019e-02	5.835e-02	6.458e-02
p-value	1.592e-05	4.126e-08	9.058e-08	3.579e-13	5.936e-05	2.488e-07
Beta	0.7014	0.6196	0.7509	0.6934	0.7029	0.6129
S.E	7.464e-02	6.632e-02	6.079e-02	5.381e-02	7.402e-02	7.014e-02
p-values	5.580e-21	9.435e-21	4.721e-35	5.456e-38	2.177e-21	2.356e-18
Gamma	-	-	-	-	-0.0287	0.0365
S.E	-	-	-	-	5.178e-02	7.997e-02
p-value	-	-	-	-	0.579	0.648

Based on the concept of a 5% significance threshold, all of the envisioned parameters for the GARCH (1, 1) model are significant, whereas the error terms are normal distribution and student-t distribution. All of the envisioned parameters are relevant for the AVGARCH (1, 1) model, whereas the error time period is regular and student-t distribution. If the error component in the GJR-GARCH (1, 1) model is normal or student-t, the envisioned parameters for Gamma are not relevant.

Table 7: The results of all volatility models for NSE 30

	GARCH (1,1)		AVGARCH (1,1)		GJR-GARCH (1,1)	
	Normal	Student-t	Normal	Student-t	Normal	Student-t
Omega	0.0716	0.0924	0.0737	0.0867	0.0697	0.0909
S.E	2.792e-02	2.927e-02	2.671e-02	2.595e-02	2.768e-02	3.081e-02
p-value	1.038e-02	1.599e-03	5.763e-03	8.282e-04	1.177e-02	3.174e-03
Alpha	0.2250	0.2982	0.2180	0.2644	0.2455	0.3116
S.E	5.194e-02	5.655e-02	3.873e-02	3.793e-02	5.866e-02	5.722e-02
p-value	1.480e-05	1.346e-07	1.830e-08	3.171e-12	2.846e-05	5.138e-08
Beta	0.7189	0.6615	0.7631	0.7260	0.7233	0.6641
S.E	6.861e-02	6.331e-02	5.343e-02	4.854e-02	6.829e-02	6.800e-02
p-values	1.092e-25	1.485e-25	2.843e-46	1.414e-50	3.234e-26	1.564e-22
Gamma	-	-	-	-	-0.0450	-0.0278
S.E	-	-	-	-	4.699e-02	6.274e-02
p-value	-	-	-	-	0.338	0.658

Based on a 5% significance level, all of the expected parameters for the GARCH (1, 1) model are significant, even if the error terms are normal distribution and student-t distribution. While the error terms for the AVGARCH (1, 1) model are normal distribution and student-t distribution, all of the expected parameters are significant. The error component in the GJR-GARCH (1, 1) model is normally distributed or student-t distribution, and the expected values for Gamma are not significant.

Table 8: AIC and BIC for all the Models

	MODEL	AIC	BIC
NSE 30	GARCH(1,1) Normal	6424.20	6441.62
	GARCH(1,1) Student-t	6227.79	6251.02
	AVGARCH(1,1) Normal	6410.60	6433.82
	AVGARCH(1,1) Student-t	**6203.43**	**6232.46**
	GJR GARCH(1,1) Normal	6426.23	6455.26
	GJR GARCH(1,1) Students-t	6230.73	6265.57
NSE India	GARCH(1,1) Normal	6245.34	6262.75
	GARCH(1,1) Student-t	5958.33	5981.33
	AVGARCH(1,1) Normal	6248.06	6271.28
	AVGARCH(1,1) Student-t	**5940.79**	**5969.82**
	GJR GARCH(1,1) Normal	6263.89	6292.92
	GJR GARCH(1,1) Students-t	5960.16	5995.00

The boldfaced quantity represents the minimum value in every group. For every stock's day by day returns, exceptional version have exceptional AIC and BIC. For exceptional datasets, the version with the minimum AIC and BIC is the same. From the AIC and BIC table, we will take a look at that the AVGARCH version has the least AIC and BIC, consequently the version AVGARCH (1,1) is chosen because it is the exceptional version amongst all of the different version.

3.8.1 GARCH Neural Network Model (Combining AVGARCH(1,1) and LSTM)

The LSTM model has been shown to recall long-term sequences, utilise data sequence history, and correctly forecast future sequences. To compute the hidden state, LSTM employs a separate function. With these already known facts and the results gained from the GARCH model variety above, the best of the GARCH model variety, AVGARCH(1,1), and the LSTM, will be blended for improved volatility predictions.

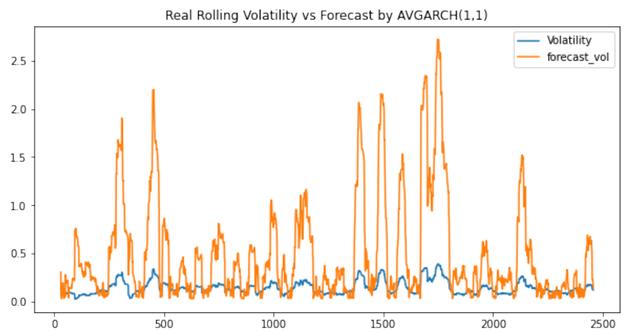

Figure 7: NSE30 Real volatility against AVGARCH forecast volatility

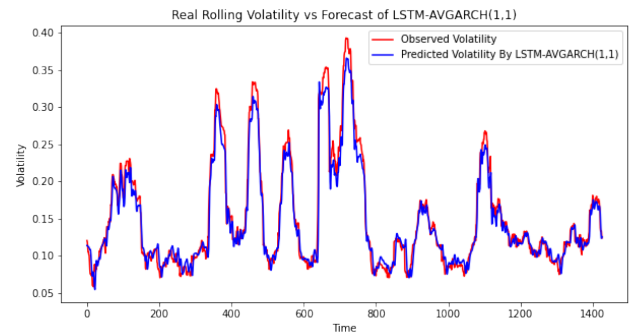

Figure 8: NSE30 Real volatility against AVGARCH +LSTM predicted volatility

The RMSE calculated from the AVGARCH model is calculated to be 0.67 while the Root Mean Squared Error of the model(AVGARCH +LSTM) is calculated as 0.093

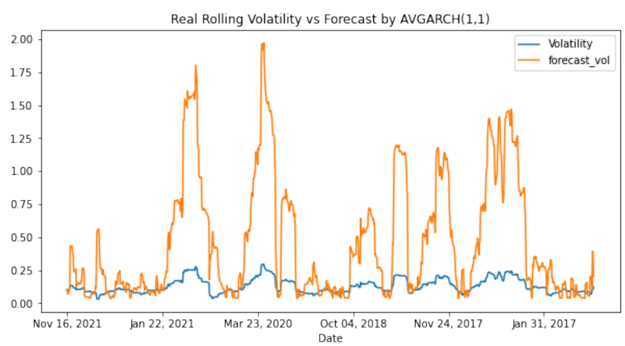

Figure 9: NSE India Real volatility against AVGARCH forecast volatility

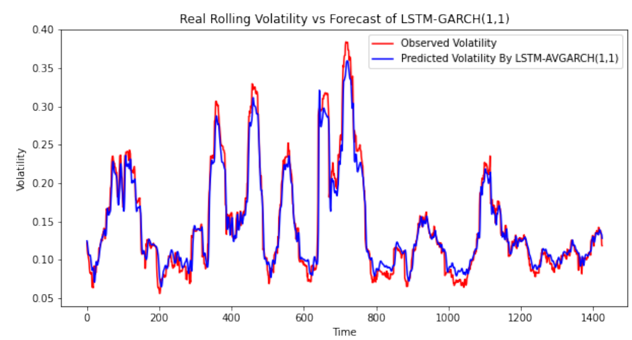

Figure 10: NSE India Real volatility against AVGARCH +LSTM predicted volatility

The RMSE calculated from the AVGARCH model is calculated to be 0.663 while the Root Mean Squared Error of the model(AVGARCH +LSTM) is calculated as 0.095
Even though AVGARCH(1,1) appears to be satisfactory for some of the exceptional types of the GARCH version, it may nevertheless be located from the figure figure 7 and figure 9 above to be a vulnerable version for time-collection facts with volatility. To enhance these, the output of the AVGARCH(1,1) changed into used as an entry for the mixed version (AVGARCH(1,1) +LSTM), creating a more potent version with the output of AVGARCH(1,1) used as input together with the actual volatility for the mixed version.

4 Conclusion

For the NSE 30 and NSE India datasets, empirical consequences confirmed that the LASSO regression version outperformed all different regression fashions used in the research. On the other hand, at the Autoregressive Conditional Heteroskedasticity fashions (ARCH), the Average Generalized Autoregressive Conditional Heteroskedasticity version (AVGARCH) beat all different fashions with the bottom Bayesian Information Criterion (BIC) and Aikanke Information Criterion. Using AVGARCH with standardized student-t determined to be the nice some of the GARCH variety, we put the AVGARCH version right into a Long Short Term Memory for the inventory volatility forecast and modelling.

References

[1] Eugene F Fama. The behavior of stock-market prices. *The journal of Business*, 38(1):34–105, 1965.

[2] Eugene F Fama. Efficient capital markets: A review of theory and empirical work. *The journal of Finance*, 25(2):383–417, 1970.

[3] Eugene F Fama. Efficient capital markets: Ii. *The journal of finance*, 46(5):1575–1617, 1991.

[4] Jeremiah Green, John RM Hand, and X Frank Zhang. The supraview of return predictive signals. *Review of Accounting Studies*, 18(3):692–730, 2013.

[5] Heiko Jacobs. What explains the dynamics of 100 anomalies? *Journal of Banking & Finance*, 57:65–85, 2015.

[6] Asher Curtis. A fundamental-analysis-based test for speculative prices. *The Accounting Review*, 87(1):121–148, 2012.

[7] Andrew W Lo and A Craig MacKinlay. Stock market prices do not follow random walks: Evidence from a simple specification test. *The review of financial studies*, 1(1):41–66, 1988.

[8] Andrew W Lo, Harry Mamaysky, and Jiang Wang. Foundations of technical analysis: Computational algorithms, statistical inference, and empirical implementation. *The journal of finance*, 55(4):1705–1765, 2000.

[9] Yaser S Abu-Mostafa and Amir F Atiya. Introduction to financial forecasting. *Applied intelligence*, 6(3):205–213, 1996.

[10] Arthur L Samuel. Some studies in machine learning using the game of checkers. ii—recent progress. *IBM Journal of research and development*, 11(6):601–617, 1967.

[11] Christopher Krauss, Xuan Anh Do, and Nicolas Huck. Deep neural networks, gradient-boosted trees, random forests: Statistical arbitrage on the s&p 500. *European Journal of Operational Research*, 259(2):689–702, 2017.

[12] Ming-Wei Hsu, Stefan Lessmann, Ming-Chien Sung, Tiejun Ma, and Johnnie EV Johnson. Bridging the divide in financial market forecasting: machine learners vs. financial economists. *Expert Systems with Applications*, 61:215–234, 2016.

[13] Michel Ballings, Dirk Van den Poel, Nathalie Hespeels, and Ruben Gryp. Evaluating multiple classifiers for stock price direction prediction. *Expert systems with Applications*, 42(20):7046–7056, 2015.

[14] MuhammadIdreesandAbdullah MatYusoff Chong, ChooWeiandAhmad. Performanceofgarchmodelsinforecasting-stockmarketvolatility. *Journalofforecasting*, 18(5):333–343, 1999.