International Journal of Statistics and Applied Mathematics 2020; 5(6): 20-27

International Journal of Statistics and Applied Mathematics

**ISSN:** 2456-1452
Maths 2020; 5(6): xx-xx

www.mathsjournal.com
Received: 11-09-2020
Accepted: 16-10-2020

**Gokul Subramaniam**
Department of Computer Science and Engineering, Kumaraguru College of Technology, Coimbatore, Tamil Nadu, India

**Indhuja Muthukumar**
Department of Statistics, PSG College of Arts and Science, Coimbatore, Tamil Nadu, India

**Corresponding Author:**
**Gokul Subramaniam**
Department of Computer Science and Engineering, Kumaraguru College of Technology, Coimbatore, Tamil Nadu, India

# Efficacy of time series forecasting (ARIMA) in post-COVID econometric analysis

**Gokul Subramaniam and Indhuja Muthukumar**

**DOI:** https://doi.org/10.22271/maths.2020.v5.i6a.609

**Abstract**
With the COVID-19 pandemic destabilizing the global economy, statisticians all over the world are striving to predict the recovery of the same. Stock market, being a reliable indicator of the performance of major economies, is usually forecasted using time series analysis methods like the Autoregressive Moving Average Model (ARIMA). In this context, We have felt the need to question the appropriateness of ARIMA in Econometric analysis, specifically in forecasting share prices in the stock market following a colossal anomaly in the general trend such as the one caused by the COVID-19 pandemic induced market instability. In this study, the predictive efficacy of the ARIMA model relative to the variations and anomalies in the data is studied. The functional constituents of the ARIMA model are also explored and criteria such as AIC (Akaike Information Criterion) and BIC (Bayesian information criterion) are employed to select the best fit model. By analyzing the share prices of twenty random stocks listed in NYSE/NASDAQ from September 2019 to August 2020, the possible correlation between predictive efficiency of the ARIMA model and variation in the data is explored and quantified.

**Keywords:** ARIMA, stock market, COVID, correlation

## 1. Introduction
For Econometric analysis, the patent and the comparatively comprehensive indicator of the performance of the economy is the stock market. Consequently, Quantitative analysis within the scope of stock market analysis has garnered the wide attention of statisticians. Its scale of impact and volatility has kept statisticians intrigued for decades, and motivated them to perpetually strive towards improving the predictive accuracy of mathematical models [11, 12].

With over $40 Trillion, NYSE and NASDAQ represent the largest stock exchanges in the world by market capitalization. Subsequently, it is no surprise that the rise or fall of share prices in these exchanges directly/indirectly affect investor sentiment, Gross Domestic Product (GDP), unemployment rates, and consumer-producer price inflation [24, 25].

In the wake of the COVID-19 pandemic, stock exchanges all over the world have plummeted to historical lows. Many sophisticated methods of forecasting the economy proved futile, as an anomaly in the stock exchange data at the scale as seen during the COVID-19 pandemic is unprecedented.

In statistics, the models used for predicting stock prices are either stochastic or deterministic in nature. Due to the randomness/white noise generated by the stock market data, stochastic models are preferred to deterministic models for effective prediction. Among the stochastic models, random walk models, especially a subclass consisting of the autoregressive group of models are considered to be a better fit for stock market forecasting [2].

The most prevalent modes of forecasting share prices in stock exchange involve time series forecasting methods such as ARMA, ARIMA, and SARIMA models. The ubiquitousness of these models owes it to its conceptual simplicity i.e. using stochastic time-series data of the past to predict the future. These models have high precision and are applicable for short term forecasting [3, 5]. Autoregressive Integrated Moving Average/ARIMA (p, d, q) models, also known as Box-Jenkins Models, are usually employed in forecasting time series data showing signs of stationarity such as share prices in stock exchanges [22].

Though the ARIMA class of models is known to be robust and effective for forecasting financial time-series data, even better than the complex Artificial Neural Network (ANN) methods, their predictive efficiency relative to the spread and anomalies in the data is seldom studied [2, 8].
This study questions the aptness of usage of ARIMA models for forecasting the economy, particularly share prices in stock exchanges, following a large scale anomaly in the general trend of data i.e. COVID-19 pandemic induced market instability. For this, twenty random stocks listed in NYSE/NASDAQ which plunged and recovered during the early weeks of the coronavirus pandemic are studied. Such stocks displaying almost V-shaped recovery are appropriate to this study since they deftly represent data that went through an anomaly and got back to its initial trend.
Using statistical measures, this paper explores the possible correlation between the variations among the time series data points (share price), to their relative effect on the predictive accuracy of the ARIMA model. This paper would hopefully serve as a preamble for further research trying to dive deep into the nuances of predictive accuracy of such autoregressive models.

## 2. Study area

Twenty stocks listed in NYSE/NASDAQ from varied sectors are considered. Like any global crisis, the COVID-19 pandemic took a toll on these stock prices. The fear of the pandemic and pandemic induced recession collapsed the US stock market during the first week of March 2020. This drove other economic indicators such as the unemployment rate to a record high (13%). For this study, the anomaly period is observed from 06/03/2020 to 28/04/2020. The easing of the lockdown and quarantine measures and increased anticipation towards an imminent vaccine has constituted an almost V-shaped recovery of most of the stocks. Prominent stock price indicator S&P 500 eclipsed its 200 day moving average while NASDAQ-100 recovered 42% from its local minimum on 23rd March. With this context established, the datasets and methods employed in time series analysis of the selected stocks can be introduced [26].

## 3. Data collection

For this research, we have obtained the consolidated daily data for the twenty selected stocks listed in NYSE/NASDAQ from Yahoo Finance website. All the sets range from September of 2019 to August of 2020. We have divided the dataset into two parts, training and test data. 231 data points (03/09/2019 - 31/07/2020) are used as training data and the rest 26 data points (01/08/2020 - 01/09/2020) are used as test data.
Each dataset has multiple fields including Date, Open, High, Low, Close, Adj Close and Volume. Among which closing price (‘Close’) of the stock is chosen to represent the share price to be predicted, as it reflects the activities of the trading day.

## 4. Materials and Methods

Monitoring the quantified effect of a phenomenon such as COVID-19 as a function of time is paramount in forecasting future responses. Since the changes in share prices in stock exchanges could be interpreted as t-dependant stochastic time series, both statistical and artificial intelligence-based methods could be employed to predict future responses [2].
The Box-Jenkins Method or the ARIMA (Auto-Regressive Integrated Moving Average) Method defines a set of models which are appropriate to forecasting the performance of stocks in the United States. The tool used for the implementation of the methods is R Studio Version 3.6.0.

### A. Preamble to ARIMA model

The ARIMA class of models attempts at defining a time series based on its own lags and lagged forecast errors, thereby aiding prediction of future values. ARIMA (p, d, q) model is an upgrade to the ARMA (p, q) model, which is an effective concoction of the linear-regressive AR (p) (Auto-Regression) and MA (q) (Moving Average) Models [23].
The primary prerequisite to employ an ARMA (p, q) model, or AR (p)/MA (q) models for that matter, is stationarity of the time series. Most of the real-life usage of the time series is non-stationary. For instance, the human population data over the past millennium show an obviously increasing trend. Thus, the stationarity test should be carried out to determine the relation between data and time. In the case of non-stationarity, commonly, We difference and detrend the data and then apply the ARMA (p, q) model for forecasting [22, 23]. This aspect is embedded within the ARIMA (p, d, q) through one of its parameters, ‘d’, which represents the order of differencing to eliminate non-stationarity.
The two main functions governing the order of the ARIMA model are Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF). In simple terms, Autocorrelation or Serial Correlation is regression of a time series onto itself. It elucidates the correlation between the present values of a time series with its past [16]. ACF plot is plotted with the correlation coefficient in the x-axis and the number of lags in the y-axis. Partial Autocorrelation (PACF), as the name suggests, explains the amount of correlation between the present value of a time series with only a particular lag of itself, with possible correlations of intervening/lower order lags removed. PACF plot is plotted with the partial correlation coefficient of lags in the x-axis and the number of lags in the y-axis [21]. The inferences required from the ACF and PACF plots for ARIMA include:

- For AR Process: The ACF plot should gradually decay, whereas the PACF plot should depict a sharp drop after ‘p’ significant lags.
- For MA Process: The ACF plot should depict a sharp drop after ‘q’ significant lags, while PACF plot shows a decreasing trend.

The ‘p’- pertains to the order of the Auto-Regressive term in the ARIMA model; it represents the number of lags to be used as predictors. This term is estimated using the PACF (Partial Auto-Correlation Function) Correlogram. The resulting AR (p) sub-model is defined as

$$y_t = c + \Phi_1 y_{t-1} + \Phi_2 y_{t-2} \dots + \Phi_p y_{t-p} + \varepsilon_t \qquad \text{(i)}$$

Where C is a constant, εt represents the white noise, $\boldsymbol{\Phi}1\ldots..\boldsymbol{\Phi}p$ are the Auto-Regression parameters and p is the Auto-Regression parameter.
The ‘q’ - pertains to the order of the Moving Average term in the ARIMA model; it represents the number of lagged errors in the forecast to be included in the model. This term is estimated using the ACF (Auto-Correlation Function) Correlogram. The resultant MA (q); the incorporation between the data and calculated residual error is defined as

$$y_t = c + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q} \qquad \text{(ii)}$$

Where C is a constant, $\varepsilon_t \ldots \varepsilon_{t-q}$ represent the white noise terms, $\theta_1 .. \theta_q$ are moving average parameters and q is the moving average order.
The integration of the above-defined fundamental sub-models results in the final form of the ARIMA model defined as

$$y'_t = c + \Phi_1 y'_{t-1} + \cdots + \Phi_p y'_{t-p} + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t \quad \text{(iii)}$$

Where p and q are Auto-Regression and Moving Average orders respectively. The fundamental models can also be accessed from the above equation by setting the respective orders to zero.

**B. Selecting the best ARIMA (p, d, q) model**
For determining the relative effectiveness of the ARIMA class of models i.e, to find the best fit model for the stock price under consideration, few criterions are examined:

- AIC (Akaike Information Criteria): AIC parameter evaluates the efficacy of each model relative to other models for a given set of data. AIC uses maximum likelihood estimation (L) a.k.a log-likelihood of a model to quantify its fitness. Though increasing the parameters maximizes the log-likelihood, it certainly leads to over fitting of the data. To balance this, AIC adds a penalty term for models with higher complexity. Lower the AIC value, the better fit is the model. The model with the lower AIC value is preferred [16, 19, 20]. AIC is calculated as follows, where it represents the maximum likelihood estimates of the model parameter, $log\, l(\hat{\theta})$ is the corresponding log-likelihood and p is the number of parameters,

$$AIC = -2log\, l(\hat{\theta}) + 2p \quad \text{(iv)}$$

- BIC (Bayesian or Schwarz Information Criterion): BIC parameter is very similar in function and approach to model selection as the AIC parameter, since both use maximum likelihood estimation (L) of a model as a measure of fit. BIC, like AIC, penalizes parameter complexity to reduce over fitting. Despite several subtle theoretical differences, their only major difference in practice is the size of the penalty; BIC penalizes model complexity more heavily [16, 19, 20]. The model with the lower BIC value is preferred. BIC is calculated as below, Where represents the maximum likelihood estimates of the model parameter, $log\, l(\hat{\theta})$ is the corresponding log-likelihood and p is the number of parameters,

$$BIC = -2log\, l(\hat{\theta}) + plog\, n \quad \text{(v)}$$

The following table summarises the observed lowest values of AIC and BIC, and their corresponding order of the chosen ARIMA model for the chosen stock indices.

**Table 1:** AIC and BIC values of selected ARIMA models

| S. No. | Company/Index Name | NYSE/NASDAQ | AIC | BIC | Order of ARIMA |
|---|---|---|---|---|---|
| 1. | S&P 500 Index | .INX | 2501.93 | 2512.14 | (0,1,2) |
| 2. | Advanced Micro Devices, Inc | AMD | 928.75 | 938.96 | (1,1,0) |
| 3. | Amazon.com, Inc | AMZN | 2438.1 | 2448.31 | (0,1,1) |
| 4. | The Boeing Company | BA | 1707.91 | 1711.33 | (0,1,0) |
| 5. | Berkshire Hathaway | BRK-A | 4614.02 | 4624.23 | (0,1,2) |
| 6. | Costco Wholesale Corporation | COST | 1390.33 | 1410.58 | (4,1,1) |
| 7. | The Walt Disney Company | DIS | 1167.39 | 1187.64 | (3,1,2) |
| 8. | Domino's Pizza, Inc | DPZ | 1650.93 | 1657.75 | (1,1,0) |
| 9. | Ford Motor Company | F | -64.25 | -60.83 | (0,1,0) |
| 10. | Peloton | PTON | 810.32 | 816.95 | (0,1,0) |
| 11. | Qualcomm | QCOM | 1059.03 | 1065.85 | (1,1,0) |
| 12. | AstraZeneca plc | AZN | 660.83 | 677.76 | (3,1,1) |
| 13. | Salesforce.com, Inc | CRM | 1298.23 | 1308.44 | (0,1,2) |
| 14. | Dell | DELL | 812.65 | 816.08 | (0,1,0) |
| 15. | Facebook, Inc | FB | 1394.49 | 1401.31 | (1,1,0) |
| 16. | The Home Depot, Inc | HD | 1431.87 | 1442.07 | (0,1,2) |
| 17. | McDonald's Corporation | MCD | 1338.48 | 1348.69 | (0,1,2) |
| 18. | NIKE, Inc | NKE | 1015.38 | 1025.6 | (1,0,0) |
| 19. | SAP SE | SAP | 1128.02 | 1144.94 | (3,1,1) |
| 20. | Nvidia Corporation | NVDA | 1648.03 | 1668.28 | (3,1,1) |

## 5. Experiments

For Forecasting, We've selected the optimal order for the ARIMA model with the help of AIC and BIC as tabulated in Table 1. Plots with the observed values of the share price with the ARIMA forecasted values superimposed is plotted using the 'Arima' function from R Studio. In the following plots, the red curve depicts the observed values of various stocks over a uniform period (Sept 2019 - Sept 2020). The Blue curve towards the end visualizes the chosen ARIMA model for that particular stock index along with its confidence interval. It is observed that the predicted values from the ARIMA model exudes a fairly linear interpretation of the data.

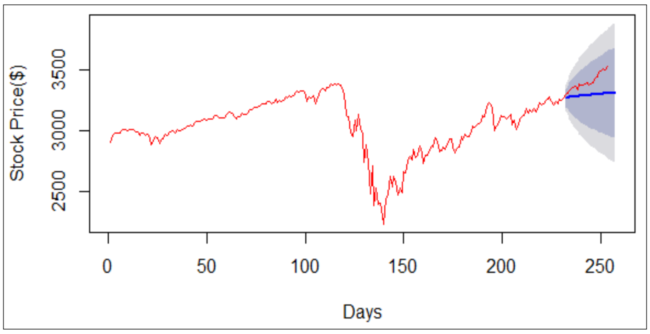


**Fig 1:** S & P 500 index (.INX) – ARIMA (0, 1, 2)

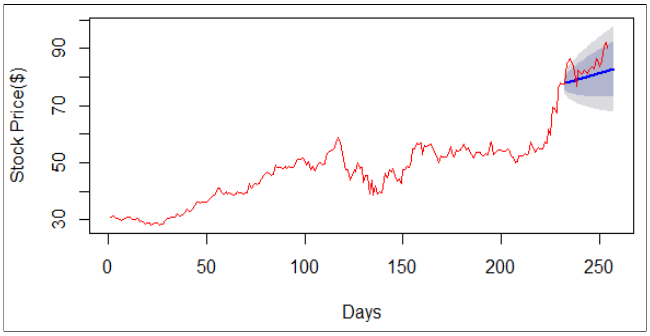


**Fig 2:** Advanced Micro Devices, Inc (AMD) – ARIMA (1, 1, 0)

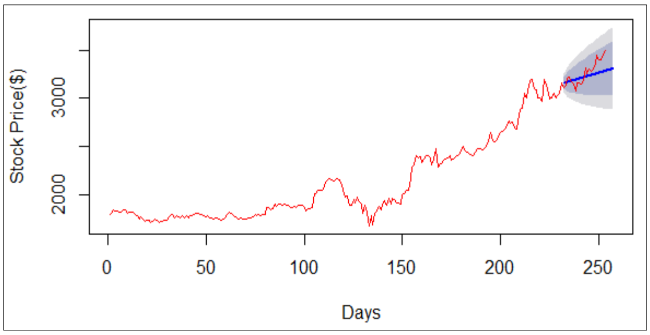


**Fig 3:** Amazon.com, INC (AMZN) – ARIMA (0, 1, 1)

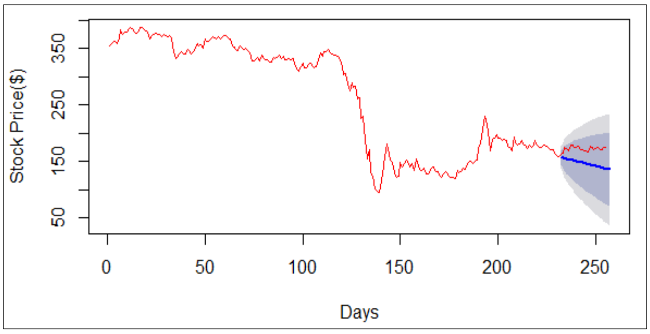


**Fig 4:** The Boeing Company (BA) – ARIMA (0, 1, 0)

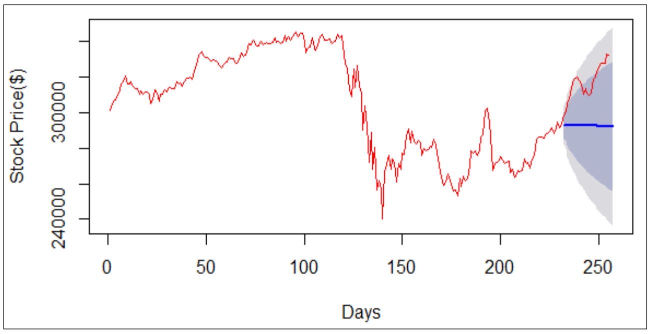


**Fig 5:** Berkshire Hathaway (BRK-A) – ARIMA (0, 1, 2)

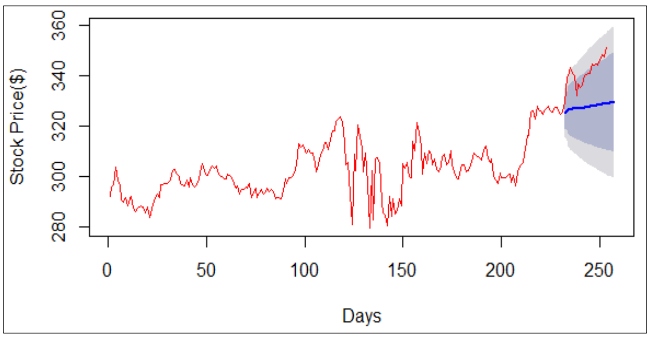


**Fig 6:** Costco Wholesale Corporation (COST) – ARIMA (4, 1, 1)

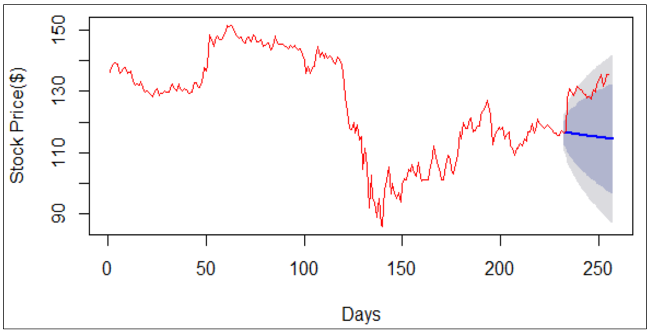


**Fig 7:** The Walt Disney Company (DIS) – ARIMA (3, 1, 2)

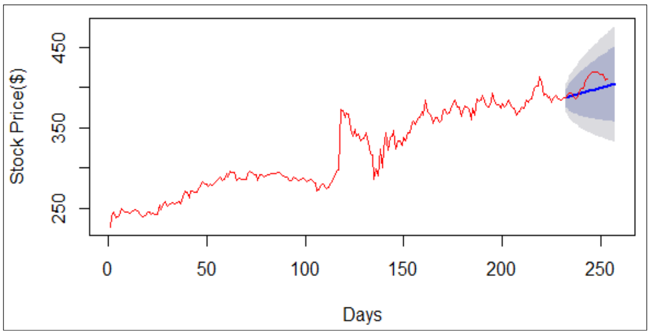


**Fig 8:** Domino's Pizza, Inc (DPZ) – ARIMA (1, 1, 0)

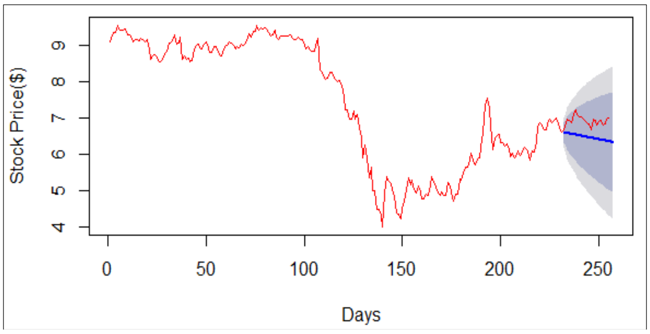


**Fig 9:** Ford Motor Company (F) – ARIMA (0, 1, 0)

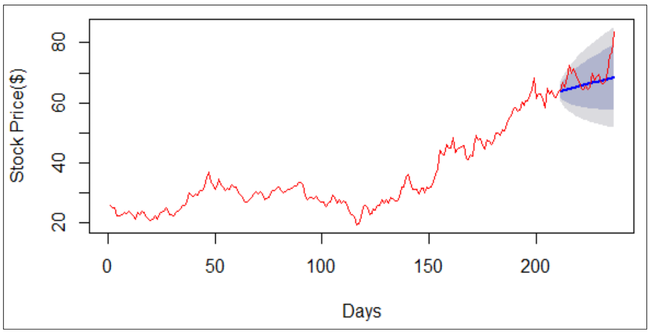


**Fig 10:** Peloton (PTON) – ARIMA (0, 1, 0)

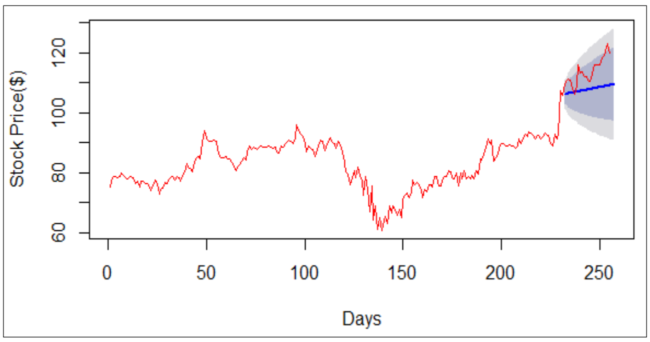


**Fig 11:** QuadComm (QCOM) – ARIMA (1, 1, 0)

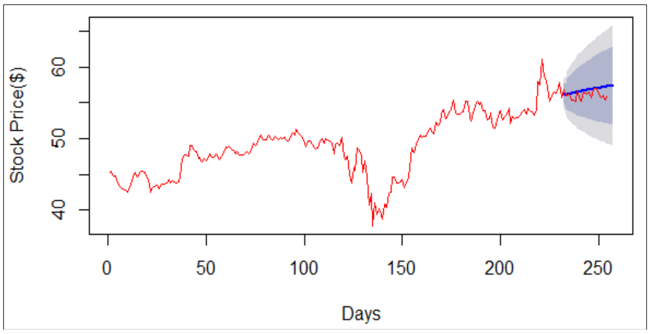


**Fig 12:** AstraZeneca plc (AZN) – ARIMA (3, 1, 1)

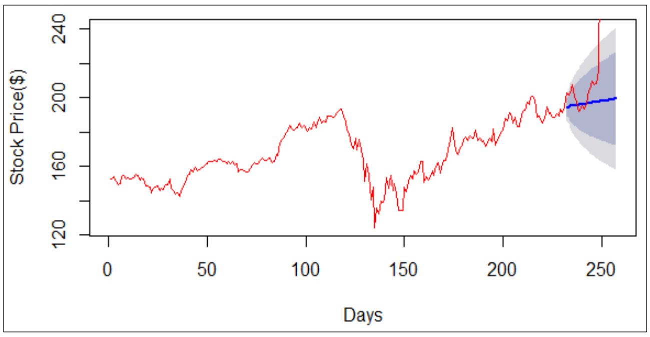


**Fig 13:** Salesforce.com, INC (CRM) – ARIMA (0, 1, 2)

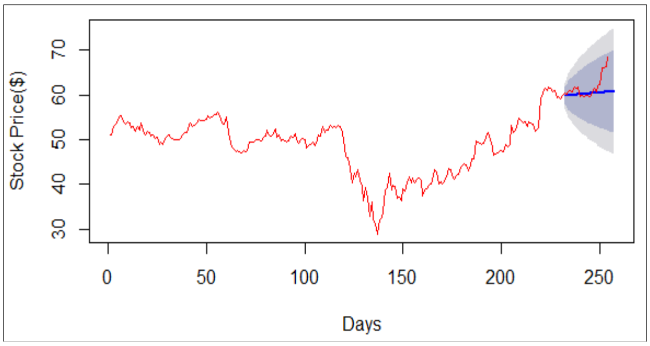


**Fig 14:** Dell (DELL) – ARIMA (0, 1, 0)

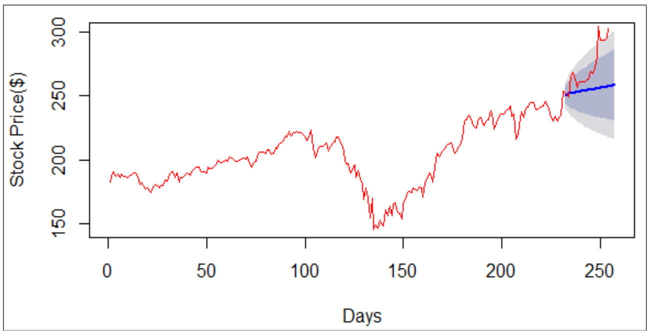


**Fig 15:** Facebook, Inc (FB) – ARIMA (1, 1, 0)

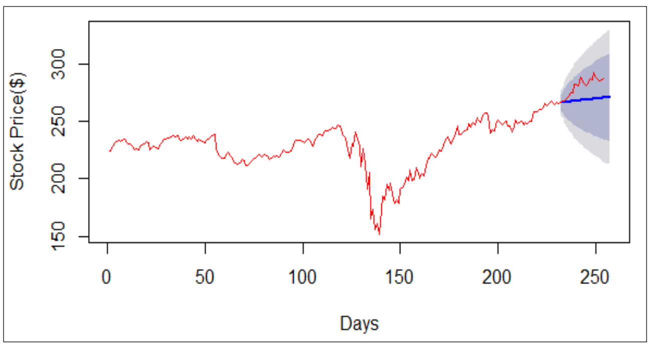


**Fig 16:** The Home Depot, Inc (HD) – ARIMA (0, 1, 2)

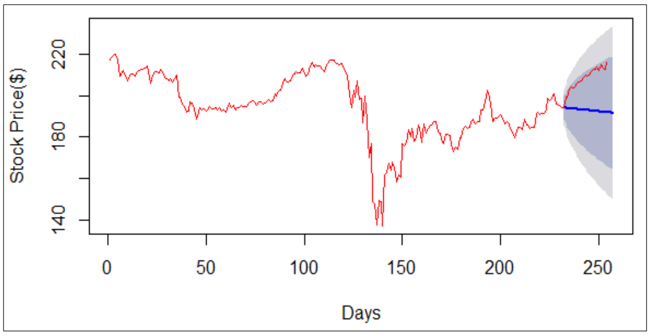


**Fig 17:** McDonald's Corporation (MCD) – ARIMA (0, 1, 2)

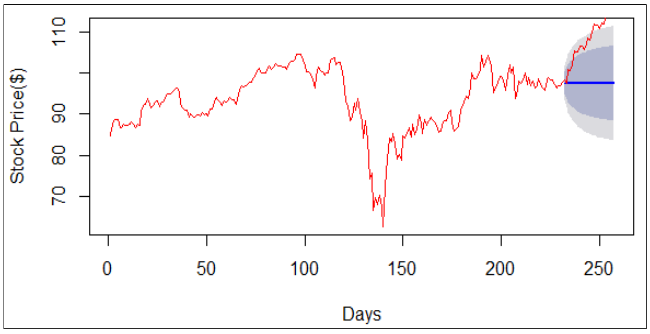


**Fig 18:** NIKE, Inc (NKE) – ARIMA (1, 0, 0)

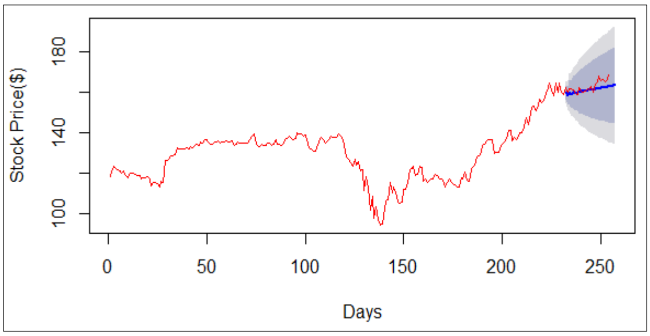


**Fig 19:** SAP SE (SAP) – ARIMA (3, 1, 1)

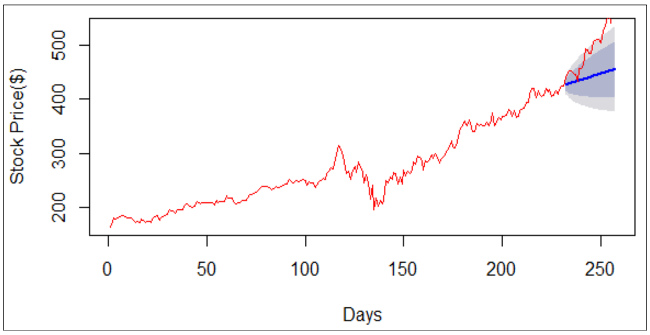


**Fig 20:** Nvidia Corporation (NVDA) – ARIMA (3, 1, 1)

## 6. Results and Discussion

From the time series analysis, various metrics are utilized to measure the performance of the chosen models. In this study, only the closing prices of the stocks are used for time series analysis using ARIMA. Relative Standard Deviation(R-SD) is used as a measure of spread in the data. About 80% of the total collected is used for training each of the ARIMA models and the rest is utilized for testing purposes.

Here, Adjusted $R^2$ is used to represent the percent of variation explained by the model i.e, measure how well the variation in the observed values is replicated by the chosen ARIMA model, and it only considers the effect of those independent variables which have an influence on the dependent variable.

The closer the value is to one, the better is the efficacy of the model in explaining the variation [15].

$$R^2_{adj} = 1 - \left[\frac{(1-R^2)(n-1)}{n-k-1}\right] \quad \text{(vi)}$$

Mean Squared Error (MSE) is a parameter which measures the average squared difference between the predicted and actual values. Here, it is employed to compute the same between the estimated values from the ARIMA model and the actual observed values.

$$\text{MSE} = \frac{1}{N}\sum_{i=1}^{N} (y_i - \hat{y}_i)^2 \quad \text{(vii)}$$

MSE is generally a measure of quality of the model, and closer the value is to zero the better. MSE helps us in determining the required quantity of RMSE, which in turn provides information on the standard deviation of the unexplained variance [13]. RMSE is calculated as below:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{N} (y_i - \hat{y})^2}{N}} \quad \text{(viii)}$$

Similarly, Mean Absolute Error (MAE), a measure of average absolute differences, is used as a loss function here. It summaries the magnitude of errors in the predictions of the ARIMA model, without considering the direction. It is calculated as follows:

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N} |y_i - \hat{y}_i| \quad \text{(ix)}$$

MAPE is derived from (vii) and (ix).

**Table 2:** Calculated values of adjusted $R^2$, RMSE, MAE and MAPE

| S. No. | NYSE/NASDAQ | Mean | SD | Adjusted $R^2$ | RMSE | MAE | MAPE (%) |
|---|---|---|---|---|---|---|---|
| 1. | .INX | 3068.8527 | 221.592 | 0.9395 | 54.7993 | 33.9749 | 1.1805 |
| 2. | AMD | 49.3964 | 10.0522 | 0.9681 | 1.7941 | 1.1581 | 2.4255 |
| 3. | AMZN | 2214.9922 | 418.1052 | 0.987 | 47.7418 | 32.4357 | 1.4869 |
| 4. | BA | 254.2686 | 97.2032 | 0.9898 | 9.8119 | 6.6615 | 3.2875 |
| 5. | BRK-A | 305623.0902 | 28320.61 | 0.9637 | 5408.213 | 3534.127 | 1.2087 |
| 6. | COST | 306.0930 | 10.7322 | 0.9395 | 4.8178 | 3.1754 | 1.0524 |
| 7. | DIS | 125.9422 | 16.7021 | 0.9685 | 2.9693 | 2.0329 | 1.7522 |
| 8. | DPZ | 325.2848 | 50.683 | 0.9712 | 8.6217 | 5.1754 | 1.5903 |
| 9. | F | 7.3874 | 1.7168 | 0.9852 | 0.2087 | 0.1452 | 2.2477 |
| 10. | PTON | 38.3289 | 14.9442 | 0.9811 | 1.6611 | 1.2179 | 3.5887 |
| 11. | QCOM | 85.6138 | 7.9417 | 0.9095 | 2.3857 | 1.6586 | 2.0581 |
| 12. | AZN | 48.8790 | 4.3274 | 0.9479 | 1.0171 | 0.7118 | 1.4726 |
| 13. | CRM | 166.7950 | 16.4895 | 0.9416 | 4.0008 | 2.7234 | 1.6733 |
| 14. | DELL | 48.5160 | 6.3442 | 0.951 | 1.4060 | 1.0002 | 2.2110 |
| 15. | FB | 201.7900 | 23.6957 | 0.9563 | 4.9482 | 3.3820 | 1.7340 |
| 16. | HD | 228.3500 | 20.6108 | 0.9332 | 5.3519 | 3.3579 | 1.5521 |
| 17. | MCD | 194.5100 | 15.6421 | 0.9234 | 4.3704 | 2.6989 | 1.4669 |
| 18. | NKE | 93.1835 | 7.5767 | 0.92 | 2.1387 | 1.4820 | 1.6703 |
| 19. | SAP | 128.8856 | 13.4732 | 0.9588 | 2.7353 | 1.8869 | 1.5093 |
| 20. | NVDA | 288.7081 | 72.0709 | 0.9862 | 8.4485 | 5.9081 | 2.2177 |

Table 2. provides the goodness of fit statistics, where the $R^2$ values (> 0.92) suggests that it could be the best fit for the model. Also there is less scope of improvement as the unexplained variabilities are under 0.08%. Furthermore, Mean Absolute Error indicates that the predicted cases are not so far from the observed cases.

To determine the spread of the training data and their supposed influence on the predictor (ARIMA (p, d, q) model) efficacy, the two parameter of Relative Standard Deviation (R-SD) and Normalized Root Mean Squared Error (NRMSE) is used.

As observed in Table 2, the mean and standard deviation of the training data is varied in scale. Relative Standard Deviation or Coefficient of Variation (cv), expressed as a percentage, is a standardised measure of dispersion of data. It is calculated as a ratio of standard deviation of the training data to the mean of the training data, multiplied by hundred. Since the values are represented in a unified scale (%), Relative Standard Deviation aids in comparing the relative spread of share prices of stocks. Since R-SD portrays how significant the deviations are as compared to the mean of the training data, larger R-SD means comparatively less variation in the data and vice versa [18]. Coefficient of Variation is calculated as below:

$$c_v = \frac{\sigma}{\mu} \quad \text{(x)}$$

Similarly, From the Table 2, it could be observed that Root Mean Squared Error (RMSE), used here as a measure of standard deviation among the residuals, is varied in scale. This need for normalization is fulfilled by dividing the calculated RMSE value with the corresponding standard deviation of the training data. It is usually multiplied by hundred and expressed as a percentage. Normalized RMSE (NRMSE) is calculated as follows,

$$\text{NRMSE} = \frac{\text{RMSE}}{\sigma} \quad \text{(xi)}$$

**Table 3:** Calculated values of R-SD and NRMSE

| S. No. | NYSE/NASDAQ | R-SD (%) | NRMSE by SD (%) |
|---|---|---|---|
| 1. | .INX | 7.22068 | 23.2688 |
| 2. | AMD | 20.35003 | 12.2943 |
| 3. | AMZN | 18.87615 | 9.2747 |

| | | | |
|---|---|---|---|
| 4. | BA | 38.22856 | 3.0891 |
| 5. | BRK-A | 9.26652 | 19.778 |
| 6. | COST | 3.50619 | 31.3006 |
| 7. | DIS | 13.26172 | 18.5603 |
| 8. | DPZ | 15.58112 | 15.2453 |
| 9. | F | 23.23959 | 12.7378 |
| 10. | PTON | 38.98934 | 10.4156 |
| 11. | QCOM | 9.27619 | 20.0542 |
| 12. | AZN | 8.85329 | 22.9189 |
| 13. | CRM | 9.88609 | 24.2863 |
| 14. | DELL | 13.07651 | 22.1708 |
| 15. | FB | 11.74275 | 20.8826 |
| 16. | HD | 9.02597 | 25.9797 |
| 17. | MCD | 8.0418 | 27.948 |
| 18. | NKE | 8.13095 | 28.2276 |
| 19. | SAP | 10.45361 | 20.3394 |
| 20. | NVDA | 24.96324 | 2.5076 |

With the measure of variation in the training data and the measure of error in the validation data normalized and tabulated as seen in Table 3, it is time to explore the possible influence of variation in the training data on the predictive accuracy of the ARIMA model.

Upon a quick look on the calculated data in Table 3, it is apparent that some inverse proportionality exists between the two variables. For Example NYSE/NASDAQ: HD with an R-SD value of 9.02597 has an NRMSE value of 0.259797, whereas NYSE/NASDAQ: BA with an R-SD value of 38.2285 has an NRMSE value of 0.03089. This conjecture could be statistically quantified using Pearson's Correlation Coefficient (PCC).

Pearson's Correlation Coefficient or Bivariate Correlation is a statistic that quantifies the linear correlation between two variables. The value of PCC lies in between +1 and -1. A negative value indicates negative correlation and vice versa. The closer the value of PCC to +1 or -1, the greater the correlation between the variables. Here, PCC is calculated for R-SD and NRMSE.

$$r = \frac{\sum (x_i - \underline{x})(y_i - \underline{y})}{\sqrt{\sum (x_i - \underline{x})^2 \sum (y_i - \underline{y})^2}} \qquad \text{(xii)}$$

Pearson's Correlation Coefficient value for R-SD and NRMSE is calculated using the above formula to be -0.8604. This signifies that there exists a strong negative correlation between the spread of the training data and the accuracy of the corresponding ARIMA (p, d, q) model (NRMSE). Moreover, this illustrates that ARIMA might not be the best choice for predicting stock prices following a large scale anomaly in the general trend of data, such as COVID-19 induces market instability, as the data is highly varied from the general trend.

## 7. Conclusion

In this study, the effectiveness of the class of ARIMA models in predicting the post COVID-19 economy (via stock market) is questioned. Twenty random stocks listed in NYSE/NASDAQ are studied for this purpose. The nature and working of the ARIMA model is studied and criteria such as AIC and BIC are used for selecting the best model for the dataset. Twelve months of share price data of these stocks are used to train the ARIMA (p, d, q) model to forecast one month of values. The prevalent correlation between the variation in the training data and the accuracy of the corresponding ARIMA model is established and quantified to be -0.8604. This signifies that, more spread or anomalous the data, the lesser is the predictive efficacy of the ARIMA model.

## 8. References

1. Saba A, Elsheikh A. Forecasting the prevalence of COVID-19 outbreak in Egypt using nonlinear autoregressive artificial neural networks. Process Safety and Environmental Protection 2020;141:1-8. Available: 10.1016/j.psep.2020.05.029
2. Pandey V, Bajpai A. Predictive Efficiency of ARIMA and ANN Models: A Case Analysis of Nifty Fifty in Indian Stock Market. International Journal of Applied Engineering Research 2019;14.
3. Mondal P, Shit L, Goswami S. Study of Effectiveness of Time Series Modeling (Arima) in Forecasting Stock Prices. International Journal of Computer Science, Engineering and Applications 2014;4(2):13-29. Available: 10.5121/ijcsea.2014.4202
4. Mustapa F, Ismail M. Modelling and forecasting S&P 500 stock prices using hybrid Arima-Garch Model. Journal of Physics: Conference Series 2019;1366:012130. Available: 10.1088/1742-6596/1366/1/012130
5. UB, SD, AP. An Effective Time Series Analysis for Stock Trend Prediction Using ARIMA Model for Nifty Midcap-50. International Journal of Data Mining & Knowledge Management Process 2013;3(1):65-78. Available: 10.5121/ijdkp.2013.3106
6. Xaba D, Moroke N, Arkaah J, Pooe C. A Comparative Study of Stock Price Forecasting Using Nonlinear Models. Risk Governance and Control: Financial Markets and Institutions 2017;7(2):7-17. Available: 10.22495/rgcv7i2art1
7. Liang J, Song W, Wang M. Stock Price Prediction Based on Procedural Neural Networks. Advances in Artificial Neural Systems 2011, P1-11. Available: 10.1155/2011/814769
8. Jarrett J, Kyper E. ARIMA Modeling with Intervention to Forecast and Analyze Chinese Stock Prices. International Journal of Engineering Business Management 2011;3: 17. Available: 10.5772/50938
9. Willmott C, Matsuura K. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. Climate Research 2005;30:79-82. Available: 10.3354/cr030079

10. Pai P, Lin C. A hybrid ARIMA and support vector machines model in stock price forecasting. Omega 2005;33(6):497-505. Available: 10.1016/j.omega.2004.07.024
11. Wang Wang J, Zhang Z, Guo S. Stock index forecasting based on a hybrid model. Omega 2012;40(6):758-766. Available: 10.1016/j.omega.2011.07.008
12. Atsalakis G, Valavanis K. Forecasting stock market short-term trends using a neuro-fuzzy based methodology. Expert Systems with Applications 2009;36(7):10696-10707. Available: 10.1016/j.eswa.2009.02.043
13. Kumar M, Thenmozhi M. Forecasting stock index returns using ARIMA-SVM, ARIMA-ANN, and ARIMA-random forest hybrid models. International Journal of Banking, Accounting and Finance 2014;5(3):284. Available: 10.1504/ijbaaf.2014.064307
14. Wang C. A comparison study between fuzzy time series model and ARIMA model for forecasting Taiwan export. Expert Systems with Applications 2011;38(8):9296-9304. Available: 10.1016/j.eswa.2011.01.015
15. Squared A. Adjusted R Squared (Meaning, Formula)| Calculate Adjusted R^2", Wall Street Mojo 2020. https://www.wallstreetmojo.com/adjusted-r-squared/
16. Adebiyi A, Adewumi A, Ayo C. Stock Price Prediction Using the ARIMA Model. 16th International Conference on Computer Modelling and Simulation 2014. Available: 10.1109/uksim.2014.67
17. Victor Devadoss A, Antony Alphonnse Ligori T. Forecasting of Stock Prices Using Multi Layer Perceptron. International Journal of Web Technology 2013;2(2):52-58. Available: 10.20894/ijwt.104.002.002.006
18. Bedeian A, Mossholder K. On the Use of the Coefficient of Variation as a Measure of Diversity. Organizational Research Methods 2000;3(3):285-297. Available: 10.1177/109442810033005
19. Brewer M, Butler A, Cooksley S. The relative performance of AIC, AICC and BIC in the presence of unobserved heterogeneity. Methods in Ecology and Evolution 2016;7(6):679-692. Available: 10.1111/2041-210x.12541
20. Vrieze S. Model selection and psychological theory: A discussion of the differences between the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). Psychological Methods 2012;17(2):228-243. Available: 10.1037/a0027127
21. Yang Q, Wang J, Ma H, Wang X. Research on COVID-19 based on ARIMA model Δ—Taking Hubei, China as an example to see the epidemic in Italy. Journal of Infection and Public Health 2020;13(10):1415-1418. Available: 10.1016/j.jiph.2020.06.019
22. Oppenheim R. Forecasting via the Box-Jenkins method. Journal of the Academy of Marketing Science 1978;6(3):206-221. Available: 10.1007/bf02729786
23. Radha S, TM. Forecasting Short Term Interest Rates Using ARMA, ARMA-GARCH and ARMA-EGARCH Models. SSRN Electronic Journal 2006. Available: 10.2139/ssrn.876556
24. Team T. NYSE May Be Bigger, But Nasdaq Is Growing Faster. Forbes 2020. https://www.forbes.com/sites/greatspeculations/2019/09/16/nyse-may-be-bigger-but-nasdaq-is-growing-faster/#2d9f77ba37c2
25. Sanger G, McConnell J. Stock Exchange Listings, Firm Value, and Security Market Efficiency: The Impact of NASDAQ. The Journal of Financial and Quantitative Analysis 1986;21(1):1. Available: 10.2307/2330987
26. Kochhar R. Unemployment rose higher in three months of COVID-19 than it did in two years of the Great Recession. Pew Research Center 2020. Available: https://www.pewresearch.org/fact-tank/2020/06/11/unemployment-rose-higher-in-three-months-of-covid-19-than-it-did-in-two-years-of-the-great-recession/