A Comparative Research of Stock Price Prediction of Selected Stock Indexes and the Stock Market by Using Arima Model

Nayab Minhaj*, Roohi Ahmed, Irum Abdul Khalique, Mohammad Imran

Department of Economics, University of Karachi, Karachi, Pakistan
E-mail: nayabminhaj@yahoo.com

Received: 25 March 2022; Revised: 27 June 2022; Accepted: 22 September 2022

Abstract: Stock prices are a really challenging and obscure task that requires tremendous efforts while the nature of the stock market is arbitrary and uncertain. Stock estimation is such an important topic in business, economics, and finance that researchers have been engaged to explain how to construct effective forecasting models. In the stock market, there is no control over the performance of an investment, so anything can occur in the short term, a pill that is difficult to swallow so researchers predict stock prices by adopting scientific methods which are valuable for investors to earn and grow their profits. In time series forecasting research, the Autoregressive Integrated Moving Average (ARIMA) models have been examined. This article explains how to use the ARIMA model to create a comprehensive stock price prediction model. The stock price of Johnson & Johnson (JNJ) is combined with published stock data from S & P (500), and a predictive model is constructed. The results demonstrate that the ARIMA model can address traditional stock price forecasting approaches and has a lot of potential for JNJ in terms of short-term forecasting. As a result of its tremendous volatility. The ARIMA model, on the other hand, is not ideal for non-stationary or weakly stationary data, such as the S & P 500 index.

Keywords: ARIMA model, stock index, dynamic forecasting, static forecasting, mean absolute error, mean absolute percentage error, theil inequality coefficient

JEL Codes: C1, C5

1. Introduction

Forecasting is a scientific subject that predicts the future levels of specific variables. The most frequent variable is demand, though other factors such as supply and pricing can also have an impact (Bozarth & Handfield, 2016). Forecasting is the technique of predicting the future values of variables under investigation (Tsay, 2000). It is also an important aspect of econometric analyses, however some variables, such as stock prices, are highly volatile and display significant drops, having to keep making prediction challenging. It is widely accepted that investing in the stock market is riskier than investing in securities or savings since securities provide a fixed return and so eliminate the uncertainty. Stock prices, on the other hand, are more volatile than the prices of other assets. Stocks, on the other contrary, can be supported to avoid risk and maximise financial gains by developing a long-term financial strategy (Atsalakis & Kimon,
Many econometrics model offerings, on the other hand, instruct us how to employ econometric tools to forecast future value.

Using an econometric model, the investor can see whether their asset will appreciate or depreciate in value in the future. Furthermore, it forecasts the future value of the existing asset, making analysis more transparent and user-friendly. Depending on the type of data, there are various techniques for economic forecasting. Single regression, exponential smoothing method, vector auto regression, ARIMA, ARCH, GARCH, and other methods are examples. Various models are organised to deal with various types of problems. A type of regression analysis known as ARIMA assesses the strength of one dependent variable in relation to other fluctuating variables. By focusing on differences between values in the series rather than actual values, the model aims to forecast future movements in the financial markets or in the value of assets. Eliminating any trends or seasonal structures is the goal of differencing. Furthermore, since Box-Jenkins established the time-series ARIMA model (Box, 1970), it has been utilised to forecast social, economic, foreign exchange, and stock concerns. Numerous scientists claim that the future values of a time series have a clear and identifiable functional relationship with the current, past, and white noise values. The foundation of the ARIMA models is the idea that past values may still have an impact on present or future values. so that when considering how much to give or accept for security, potential buyers and sellers will presume that recent market transactions have some influence on the stock. Although this presumption holds true in many other situations, such as a shock from the outside. According to Ho and Xie (1998), for consistency forecasting and evaluation, ARIMA models are used. The ARIMA model has undoubtedly aided dependability specialists in quantifying the dynamic characteristics of assets. For the investigation of reliability evaluation, the ARIMA modelling approach is also a viable alternative. In their study, Fattah et al. (2018) analysed historical data to support the ARIMA model for predicting. This model can be used to estimate and anticipate future demand in the food manufacturing industry, according to the findings. The findings would provide sound advice for making decisions. The ARIMA model has been used in a wide range of different fields. In their study, Al Wadi et al. (2010) employed the ARIMA model to predict the closing price dataset. Al Wadia and Ismail (2011) also generated a financial time series dataset that could be useful. The ARIMA model was utilised by Al Wadi et al. (2013) to forecast insurance time series models. In his research for forecasting financial time series data, (Alwadi, 2015) introduces the ARIMA model to meet real-world difficulties. To project financial data, the ARIMA model is utilised, which is suitable. As a consequence of the analysis, the ARIMA model provided a suitable future proposal, and his research recommended that the ARIMA model be used.

By employing the Autoregressive Integrated Moving-Average (ARIMA) or Autoregressive Moving-Average (ARMA) model, the ARIMA procedure analyses and forecasts equally spaced univariate time series data, transfer function data (generally all possible input for an output), and intervention data (a comparison point in analysis). As a linear mixture of its own previous values, historical errors (sometimes called shocks or innovations), and current and past values of other time series, an ARIMA model predicts as well as compares a value in a response time series. Despite the fact that Box and Jenkins popularised the ARIMA technique, ARIMA models are frequently referred to as Box-Jenkins models (Box & Jenkins, 1976). The ARIMA process is simple to use and provides a comprehensive set of tools for time series model evaluation, parameter estimation, and projection. It also enables for the evaluation of a wide range of ARIMA and ARIMAX models. The ARIMA process can be used to create a seasonal, significant part, and factored ARIMA models, intrustion or interrupted time series models, multiple regression analysis with ARMA errors, and rational transfer function models of unlimited complexity. PROC ARIMA is based on the Box-Jenkins strategy for time series modelling, and includes features for the Box-Jenkins method’s identification, estimate, diagnostic checking, and forecasting processes. The ARIMA model, according to (Mahir & Al-khazaleh, 2008), is the most general form of predicting because no assumptions are required. Furthermore, ARIMA models can be fitted to any set of time series data by calculating the parameters p, d, and q to be acceptable with the desired dataset. The ARIMA time series model family is sophisticated and strong, and its implementation demands extensive understanding. Financial market volatility and unpredictability have grown increasingly important for risk control in recent years. The standard deviation of the daily compound yield is used to calculate volatility concerns. In this analysis, Jaber et al. (2017) projected volatility in the Jordanian banking system following the 2006 crisis. ARIMA and ARIMA-wavelet are used to estimate the parameters p, d, and q. The models are then compared using a variety of accuracy metrics (Wooldridge, 2009). The results imply that ARIMA-wavelet outperforms ARIMA in terms of accuracy.
1.1 **Objective of the study**

This paper reviews the approach to forecasting based on the construction of the ARIMA time series model of a single stock and stock market overall. Its goal is to increase the amount of knowledge regarding time series forecasting and its limitation. Moreover, consider the use of such models both for the analysis of a single time series and the exploitation of relationships among series, in the production of a superior forecast.

1.2 **Contribution to the study**

This research looked at the progress made in attempting to establish forecasting of stock prices of single indices and the stock market also by using the ARIMA model to determine the efficiency of a linear model so that beginners can understand the difference between how efficiently the ARIMA model could still perform forecasting in the case of single indices and if it would be able to capture highly volatile markets. although Stock price forecasting has been the subject of numerous investigations. However, this research compared the performance of forecasting of a single stock index to the entire stock market in order to determine the ARIMA model’s capacity to forecast stock assets and other volatile variables.

1.3 **Domains for additional research**

This research suggests how well the ARIMA model could be used to forecast stock prices. Even though the ARIMA model has the ability to predict stock assets and can be effectively used in financial variable forecasting. In the future, researchers may compare the results of machine learning models like SVR, RVM, DT, and ELM to ARIMA’s results to attain the same high accuracy objective. Furthermore, as highly volatile parameters cannot be estimated using a linear model such as ARIMA or others, researchers must use other alternative statistical approaches to achieve high accuracy.

2. **Data and methodology**

The ARIMA model for stock price forecasting is developed using the following methods. This research is subdivided into several parts that include not just the construction of an ARIMA model as well as the validation of its viability for predictive analysis. For the implementation, EViews software version 10 has been used. Historical daily stock prices from the United States have been used in this research. yahoo finance is the source of the data. The stock data for Johnson & Johnson used in this analysis ranges from April 29, 2016 to April 26, 2019, with a total of 753 observations. And, the stock data of S & P 500 index used in this study covered the period from 21 April, 2014 to 17 April, 2019 with a total of 1,258 observations. Essentially, the data is comprised of five components. We use data from the open, low, high, close, and adjusted closing prices in our analysis. The closing price is used in this study to represent the price of the index to be forecasted. Closing prices were chosen since they reflect all of the index’s actions during a trading day. To reduce data noise, we employ a log transformation of the closing price, which is labelled P on the working file.

For time series forecasting, the ARIMA model is seen to be a prominent and commonly used statistical tool. Considering the concept that ARIMA stands for Autoregressive Integrated Moving Average. It’s a class of models that can capture a wide range of different periodic structures in time series data (Asteriou & Hall, 2007). The stages of the ARIMA MODEL are as follows:

- **Stage 1-**If a time series has a trend or seasonality component, the first step is to ensure that the data is stationary so that the model may be used for forecasting. We can’t proceed to the next level if the data isn’t stationary. As a result, the initial stage is obligatory.

- **Stage 2-**If the time series is not stationary in the first phase, it must be differentiated to become stationary in the second. So, first check for stationarity, then take the first difference. If the series becomes stationary after applying the first difference, we go to the next stage; if not, we apply as many differences as necessary until the series becomes stationary. Check for seasonal variations as well.

- **Stage 3-**Then make an attempt A validation sample should be filtered out. It’s used to assess if our model is...
accurate. This can be accomplished by using a train-test-validation cycle.

- Stage 4-The AR and MA criteria should now be specified. Use the ACF and PACF to determine whether AR, MA, or both terms should be included.
- Stage 5-Build the model after specifying AR and MA. Configure the model and set the number of forecasting periods to N. It is usually determined by your requirements.
- Stage 6-In the validation sample, contrast the predicted values to the actuals. As a result, the validated model is adopted to anticipate potential value.

We would go on to forecasting or predicting future values of a variable after executing those basic steps. Economic decision makers, on the other hand, frequently have access to a variety of forecasts, and then evaluating the quality of a projection necessitates comparing predicted values to actual target values across a forecast period. Setting aside some history of your actual data for comparison is a standard approach. Using EViews to evaluate ARIMA models, we may first build a forecast assessment statistic to provide a measure of prediction accuracy, and then perform Combination testing to verify if a composite average of forecasts outperforms individual forecasts.

The following criteria are utilised in this study for each stock index to find the optimum ARIMA model among multiple experiments performed. First, we look for a BIC (Bayesian or Schwarz Information Criterion) that is comparatively small, then we look for a Comparatively low sigma square and a Relatively small standard error of the regression. Furthermore, a high adjusted R2 is preferred, and then Q-statistics and a correlogram reveal that there is no significant pattern remaining in the Autocorrelation Functions (ACFs) and Partial Autocorrelation Functions (PACFs) of the residuals, implying that the selected model’s residual is white noise.

3. Results and discussion
3.1 Arima (p, d, q) model for JNJ stock index

The stock data for Johnson & Johnson used in this analysis ranges from April 29, 2016 to April 26, 2019, with a total of 753 observations. By taking the log of the closing price (p), Figure 1 illustrates the random walking trend, which suggests that the data is not stationary at the level.

![Graphical representation of the JNJ stock closing price (p) index](image-url)
The time series correlogram is shown in Figure 2. The ACF dies away quite slowly in the graph, indicating that the time series is nonstationary. By differences, a non-stationary series is transformed into a stationary series.

![Correlogram of JNJ stock price index](image)

**Figure 2.** The correlogram of JNJ stock price index

![Graphical representation of the JNJ stock price index after differencing](image)

**Figure 3.** Graphical representation of the JNJ stock price index after differencing

The series “DP” of the JNJ index becomes stationary after the first difference, as shown in Figure 3 of the line.
The series “DP” of the JNJ index becomes stationary after the first difference, as shown in Figure 4 in correlogram also.

The model was subsequently assessed in Figure 5 by using the Augmented Dickey Fuller (ADF) unit root test on the “DP” of the JNJ stock index. After the first-difference of the series, the outcome verifies that the series seems to become stationary.

Table 1. Statistical results of different Arima parameters for JNJ stock index

<table>
<thead>
<tr>
<th>ARIMA</th>
<th>BIC</th>
<th>Adjusted R2</th>
<th>Sigma square</th>
<th>S.E. of regression</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR (3), AR (4), AR (7),</td>
<td>-6.288871</td>
<td>0.021737</td>
<td>9.86E-05</td>
<td>0.010005</td>
</tr>
<tr>
<td>AR (25), MA (3), MA (4), MA (5), MA (7), MA (31)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AR (3), AR (25), MA (3), MA (5), MA (31)</td>
<td>-6.312583</td>
<td>0.015848</td>
<td>9.98E-05</td>
<td>0.010035</td>
</tr>
<tr>
<td>AR (3), MA (3), MA (5)</td>
<td>-6.319552</td>
<td>0.007470</td>
<td>0.000101</td>
<td>0.010078</td>
</tr>
<tr>
<td>AR (3), AR (7), AR (25), MA (3), MA (5), MA (7), MA (25)</td>
<td>-6.301926</td>
<td>0.019919</td>
<td>9.91E-05</td>
<td>0.010014</td>
</tr>
<tr>
<td>AR (3), AR (25), MA (3), MA (5), MA (7)</td>
<td>-6.316384</td>
<td>0.019328</td>
<td>9.94E-05</td>
<td>0.010017</td>
</tr>
</tbody>
</table>

Note: The bold row represents the best ARIMA model among the several experiments
Data source: Johnson & Johnson (JNJ) NYSE-NYSE Closing Price. Currency in USD, 2019
Figure 5. ADF unit root test for DP of JNJ stock index

Table 1 presents statistical findings for various Arima Parameters for the JNJ Stock Index.

Table 1 summarizes the various parameters of the autoregressive (p) and moving average (q) in the ARIMA models that have been investigated. For the JNJ stock index, ARIMA (AR (3), MA (3), MA (5)) is the optimum. As illustrated in Figure 6, the model returned the least Bayesian or Schwarz information criterion of -6.319552 and the smallest sigma square of 0.000101.

The output of the ARIMA ((AR (3), MA (3), MA (5)) estimation is shown in Figure 6. Figure 7 illustrates the residual pattern. According to the research, the residuals (difference between actual and predicted values) are a sequence of random errors if the model is acceptable. As a result, the proposed ARIMA model’s residual is white noise, and the time series exhibits no other noticeable features owing to the absence of notable ACF and PACF spikes. As a result, AR (p) and MA (q) are no longer relevant concerns.

Figure 8 illustrates that the probability of a square residual correlogram is greater than 5%, indicating that autocorrelation doesn’t really exists.

Even though the residual graph in Figure 9 of JNJ reveals a stationary trend, however we could still verify if heteroskedasticity is present or not by performing a residual diagnostic, ARCH test.

Figure 10 shows that the probability value is more than 5%, indicating that the model is not heteroskedastic. Because the JNJ data ARIMA model AR (3) MA (3) MA (5) is white noise, it fits the data reasonably, and we may accept this fit and move on to the next phase, forecasting.
Dependent Variable: DXP
Method: ARMA Maximum Likelihood (OPG - BHHH)
Date: 04/27/19 Time: 21.25
Included observations: 753
Convergence achieved after 32 iterations
Coefficient covariance computed using outer product of gradients

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>t-Statistic</th>
<th>Prob</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>0.060209</td>
<td>0.000511</td>
<td>0.684433</td>
<td>0.5591</td>
</tr>
<tr>
<td>AR(3)</td>
<td>0.335323</td>
<td>0.255723</td>
<td>1.323004</td>
<td>0.1862</td>
</tr>
<tr>
<td>MA(3)</td>
<td>-0.255770</td>
<td>0.269992</td>
<td>-0.947326</td>
<td>0.3438</td>
</tr>
<tr>
<td>MA(5)</td>
<td>0.077503</td>
<td>0.035174</td>
<td>2.146056</td>
<td>0.0271</td>
</tr>
<tr>
<td>SIGMASQ</td>
<td>0.006101</td>
<td>1.64E-05</td>
<td>0.0000</td>
<td></td>
</tr>
</tbody>
</table>

R-squared: 0.12750
Adjusted R-squared: 0.007470
S.E. of regression: 0.10078
Sum squared resid: 0.075669
Log likelihood: 2393.672
Hannan-Quinn criter.: -6.330423
F-statistic: 2.414663
Durbin-Watson stat: 1.584069
Prob(F-statistic): 0.047518

Inverted AR Roots: .70
Inverted MA Roots: .53 + .21i

Figure 6. ARIMA ((AR (3), MA (3), MA (5)) estimation output with DP of JNJ index

---

Date: 04/28/19 Time: 10:09
Sample: 4/28/2016 4/24/2020
Included observations: 753

<table>
<thead>
<tr>
<th>Series</th>
<th>Autocorrelation</th>
<th>Partial Correlation</th>
<th>AC</th>
<th>PAC</th>
<th>Q-Stat</th>
<th>Prob</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 7. Correlogram of residuals of the JNJ stock index
3.1.1 Forecasting of ARIMA AR (3), MA (3), MA (5)

The optimal model selected can be expressed as follows in forecasting form:
\[ P_t = a + \beta_2 P_{t-3} + \beta_1 E_{t-3} + \beta_2 E_{t-5} + e_t \] (1)

Where \( E_t = P_t - \hat{P}_t \) (i.e., the difference between the actual value of the series and the forecast value).

Figure 10. Test of residual diagnostic

![Heteroskedasticity Test: ARCH](image)

Figure 11. graph of P (actual sample) and PF (estimated sample)
Table 2 displays the empirical results of actual and predicted ARIMA AR (3) MA (3) MA (5) sample values for the JNJ stock index.

The P (actual sample) and PF (estimated sample) graphs for the JNJ stock are shown in Figure 11. The red line, which represents the actual sample, depicts the divergence in actual values from the blue line, which is estimated.

<table>
<thead>
<tr>
<th>Sample Period</th>
<th>P</th>
<th>PF</th>
</tr>
</thead>
<tbody>
<tr>
<td>4/01/2019</td>
<td>4.934330</td>
<td>4.943743</td>
</tr>
<tr>
<td>4/02/2019</td>
<td>4.925150</td>
<td>4.944042</td>
</tr>
<tr>
<td>4/03/2019</td>
<td>4.921221</td>
<td>4.944341</td>
</tr>
<tr>
<td>4/04/2019</td>
<td>4.909488</td>
<td>4.944640</td>
</tr>
<tr>
<td>4/05/2019</td>
<td>4.913977</td>
<td>4.944939</td>
</tr>
<tr>
<td>4/08/2019</td>
<td>4.913684</td>
<td>4.945238</td>
</tr>
<tr>
<td>4/09/2019</td>
<td>4.909488</td>
<td>4.945537</td>
</tr>
<tr>
<td>4/10/2019</td>
<td>4.909652</td>
<td>4.945836</td>
</tr>
<tr>
<td>4/11/2019</td>
<td>4.906829</td>
<td>4.946135</td>
</tr>
<tr>
<td>4/12/2019</td>
<td>4.912508</td>
<td>4.946434</td>
</tr>
<tr>
<td>4/15/2019</td>
<td>4.391674</td>
<td>4.946733</td>
</tr>
<tr>
<td>4/16/2019</td>
<td>4.927399</td>
<td>4.947031</td>
</tr>
<tr>
<td>4/17/2019</td>
<td>4.931015</td>
<td>4.947330</td>
</tr>
<tr>
<td>4/18/2019</td>
<td>4.923769</td>
<td>4.947629</td>
</tr>
<tr>
<td>4/22/2019</td>
<td>4.926021</td>
<td>4.947928</td>
</tr>
<tr>
<td>4/23/2019</td>
<td>4.940928</td>
<td>4.948227</td>
</tr>
<tr>
<td>4/24/2019</td>
<td>4.935912</td>
<td>4.948526</td>
</tr>
<tr>
<td>4/25/2019</td>
<td>4.939139</td>
<td>4.948825</td>
</tr>
<tr>
<td>4/26/2019</td>
<td>4.944424</td>
<td>4.949124</td>
</tr>
<tr>
<td>4/29/2019</td>
<td>NA</td>
<td>4.949423</td>
</tr>
<tr>
<td>4/30/2019</td>
<td>NA</td>
<td>4.949722</td>
</tr>
<tr>
<td>5/01/2019</td>
<td>NA</td>
<td>4.150021</td>
</tr>
</tbody>
</table>

The predicted value is close to the sample value, as shown in Figures 12 and 13, and the root means squared error is 0.05 for dynamic forecasting and 0.01 for static forecasting, which is the smallest and best for forecasting and indicates a better model fit. The mean absolute error, which measures the difference between true and predicted values, is also nearly zero in dynamics and static, measuring 0.046 and 0.006, respectively. Additionally, mean absolute percentage error, which measures forecasting error, is frequently employed for accuracy. When the high inaccuracy is
extremely undesired, its value should be less than 10%, which is considered to be very good and useful. The results show that the mean absolute error in dynamic is 0.95 and in static is 0.13, both of which are less than 10%, making them very low values that are desirable and considered to be good for forecasting accuracy. Furthermore, the Theil inequality coefficient, which ranges from 0 to 1, measures economic inequality. As a result, the Theil inequality coefficient in JNJ data is around 0.005 in dynamic and 0.001 in static, which is close to 0, indicating the superior fit.

Figure 12. Graph of P dynamic (closing price actual) and PF (forecast price) of the JNJ stock index

Figure 13. Graph of P static (closing price actual) and PF (forecast price) of the JNJ stock index
3.2 ARIMA (p, d, q) model for S & P 500 index

This study also employed stock data from the S & P 500 index, which encompassed the period from April 21, 2014 to April 17, 2019, with a total of 1,258 observations. The overall sequence of the series is shown in Figure 14. The graph demonstrates that the index has been moving steadily since 2016 and has been moving downhill since 2018.

![Graphical representation of the SNP stock index closing price (p) index](image1)

Data source: S & P 500 (^GSPC) SNP-SNP closing price. Currency in USD, 2019

**Figure 14.** Graphical representation of the SNP stock index closing price (p) index

![The correlogram of S & P 500 index](image2)

**Figure 15.** The correlogram of S & P 500 index
**Figure 16.** Graphical representation of the S & P 500 index after differencing

**Figure 17.** The correlogram of S & P 500 index after first differencing
The correlogram of the S & P 500 index time series is shown in Figure 15. The ACF dies away quite slowly in the graph, indicating that the time series is nonstationary. By differencing, a non-stationary series is turned into a stationary series. The series “DP” of the S & P 500 index becomes stationary after the first difference, as shown in Figures 16 and 17 of the line graph and correlogram, respectively.

The model is verified using the Augmented Dickey Fuller (ADF) unit root test on the “DP” of the S & P 500 index in Figure 18. After the first-difference of the series, the outcome verifies that the series becomes stationary.

![Figure 18. ADF unit root test for DP of S & P 500 index](image)

<table>
<thead>
<tr>
<th>Table 3. Statistical results of different Arima parameters for S &amp; P index</th>
</tr>
</thead>
<tbody>
<tr>
<td>Differenced adj Closing price</td>
</tr>
<tr>
<td>-----------------------------</td>
</tr>
<tr>
<td>Sigma sq Volatility</td>
</tr>
<tr>
<td>Adj R sq</td>
</tr>
<tr>
<td>Akaike info criterion</td>
</tr>
<tr>
<td>Schwarz criterion</td>
</tr>
</tbody>
</table>

Note: The bold row represents the best ARIMA model among the several experiments
Data source: S & P 500 (^GSPC) SNP-SNP Closing Price. Currency in USD, 2019
Table 3 displays the various parameters of the autoregressive (p) and moving average (q) in the ARIMA models that have been tested. For the S & P 500 index, ARIMA (8, 1, 4) is considered the best among all.

The optimal model for the S & P 500 index outlined is ARIMA (8, 1, 4). As shown in Figure 19, the model produced the least Bayesian or Schwarz information criterion of -6.715834 and the lowest standard error of regression of 0.008341.

![Figure 19](image-url) ARIMA (8, 1, 4) estimation output with D(P) of S & P 500 index

![Figure 20](image-url) Correlogram of residuals of the S & P 500 index
The series’ residual can be observed in Figure 20. An acceptable approximation’s residuals (the difference between the actual and anticipated values) are a series of random errors. The residual of the selected ARIMA model is white noise, and there are no other remarkable patterns in the time series because there are no notable ACFs and PACFs spikes. As a result, neither AR (p) nor MA (q) need to be considered anymore.

Figure 21 shows residuals in the ACF and PACF that appear to be uncorrelated. As a result, the requirement appears to be adequate. Furthermore, while there is no noticeable spike in the q statistic as shown in the figure, the correlogram of squared residual probability is less than 5%, indicating that there is an autocorrelation problem. This may be due to the fact that the mean is constant but the variance is not, necessitating the verification of heteroskedasticity.

![Figure 21. Correlogram of residuals of the S & P 500 index](image)

![Figure 22. Test of heteroskedasticity of ARIMA (8, 1, 4) of S & P 500 index](image)
Figure 22 depicts that the probability is less than 5%, indicating the presence of heteroskedasticity. As a result, the model no longer emits white noise, which is unacceptable.

4. Conclusion

This comparison between the forecast of individual stock indexes and the stock market by using a linear model paints a clear picture of the effectiveness of the ARIMA model in both situations.

4.1 ARIMA (p, d, q) model for JNJ stock index

The current closing price of JNJ stock depends on previous shocks of 3 months and the average volatility of the current month depends on volatility in the preceding 3 months and 5 months. The adjusted R-squared exhibits that previous period variation explains by 0.75% of today’s stock prices of JNJ stock. Furthermore, our ARIMA model AR (3) MA (3) MA (5) is white noise (zero mean and constant variance), so this model is healthy for forecasting in the short run.

4.2 ARIMA (p, d, q) model for the S & P 500 index

The current closing price of the S & P 500 index depends on the former shock of 8 months and current volatility depends on the volatility of the preceding 4 months. But this model (8, 1, 4) is not white noise (mean zero and constant variance) so we cannot admit a particular fit. Besides, we must start over as the BJ methodology is iterative manner. Unless moving to ARCH GARCH or any other prediction process.

The forecasting of stock prices by applying the ARIMA model explicates that the ARIMA model has the potential to foretell stock assets and can be utilized in the forecasting of financial variable efficiently in the short run. Furthermore, the ARIMA model helps investors, government regulators, policymakers and relevant stakeholders to take an informed decision. On the other hand, the giant volatile variables cannot estimate through a linear model like ARIMA and others to capture the volatility we have to propel the ARCH GARCH family or any other forecasting tools.

5. Recommendations

1. Demand projections are crucial in stock markets and other supply chains. It is one of the most important planning techniques for a company or an individual can apply in the future because of its connection to other corporate strategies.

2. The acquired results show that this model may be used to estimate and forecast future demand in stock indices, and that these outcomes would furnish investors with trustworthy guidance for making decisions.

3. To provide credible forecasts and enhance forecast accuracy in the future, it really should continue to develop new models that integrate qualitative and quantitative techniques in the future to achieve the same high accuracy target.

Conflict of interest

The authors declare that he has no conflict of interest.

References


