Global Economics Science
http://ojs.wiserpub.com/index.php/GES/

Research Article

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,

DOI: https://doi.org/10.37256/ges.4120231426

2009). 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, intrusion 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 21st April, 2014 to 17th 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.

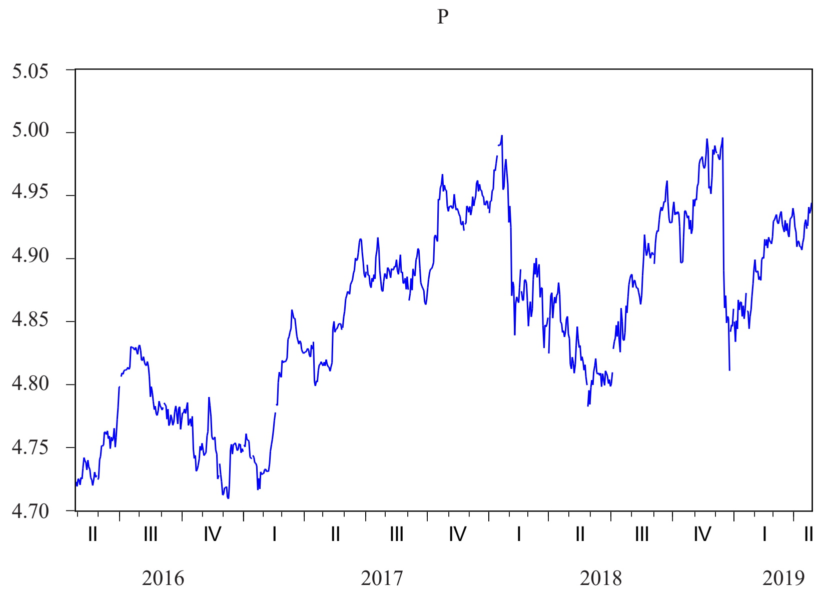
We would go on to forecasting or predicting future values of a variable after executing those basic steps. Economic decision makers, on the other hand, frequently have access to a variety of forecasts, and then evaluating the quality of a projection necessitates comparing predicted values to actual target 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 disccusion

3.1 *Arima (p, d, q) model for JNJ stock index*

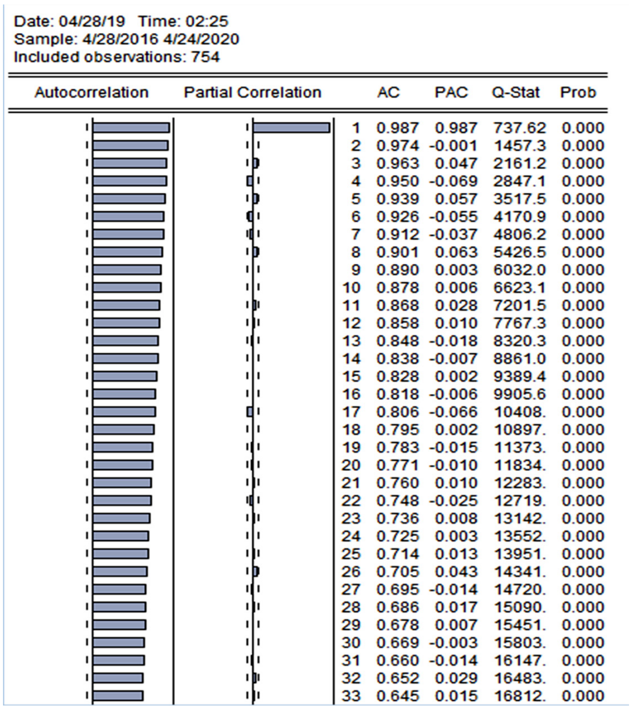
The stock data for Johnson & Johnson used in this analysis ranges from April 29, 2016 to April 26, 2019, with a total of 753 observations. By taking the log of the closing price (p), Figure 1 illustrates the random walking trend, which suggests that the data is not stationary at the level.



Data source: Johnson & Johnson (JNJ) NYSE-NYSE closing price. Currency in USD, 2019

Figure 1. Graphical representation of the JNJ stock closing price (p) index

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.



Date: 04/28/19 Time: 02:25
Sample: 4/28/2016 4/24/2020
Included observations: 754

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.987	0.987	737.62	0.000
		2	0.974	-0.001	1457.3	0.000
		3	0.963	0.047	2161.2	0.000
		4	0.950	-0.069	2847.1	0.000
		5	0.939	0.057	3517.5	0.000
		6	0.926	-0.055	4170.9	0.000
		7	0.912	-0.037	4806.2	0.000
		8	0.901	0.063	5426.5	0.000
		9	0.890	0.003	6032.0	0.000
		10	0.878	0.006	6623.1	0.000
		11	0.868	0.028	7201.5	0.000
		12	0.858	0.010	7767.3	0.000
		13	0.848	-0.018	8320.3	0.000
		14	0.838	-0.007	8861.0	0.000
		15	0.828	0.002	9389.4	0.000
		16	0.818	-0.006	9905.6	0.000
		17	0.806	-0.066	10408.	0.000
		18	0.795	0.002	10897.	0.000
		19	0.783	-0.015	11373.	0.000
		20	0.771	-0.010	11834.	0.000
		21	0.760	0.010	12283.	0.000
		22	0.748	-0.025	12719.	0.000
		23	0.736	0.008	13142.	0.000
		24	0.725	0.003	13552.	0.000
		25	0.714	0.013	13951.	0.000
		26	0.705	0.043	14341.	0.000
		27	0.695	-0.014	14720.	0.000
		28	0.686	0.017	15090.	0.000
		29	0.678	0.007	15451.	0.000
		30	0.669	-0.003	15803.	0.000
		31	0.660	-0.014	16147.	0.000
		32	0.652	0.029	16483.	0.000
		33	0.645	0.015	16812.	0.000

Figure 2. The correlogram of JNJ stock price index

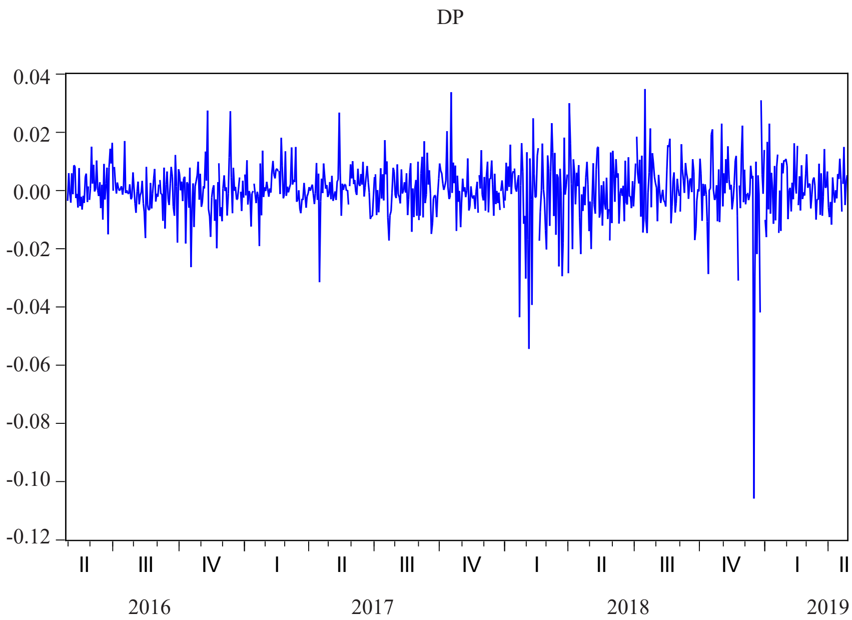
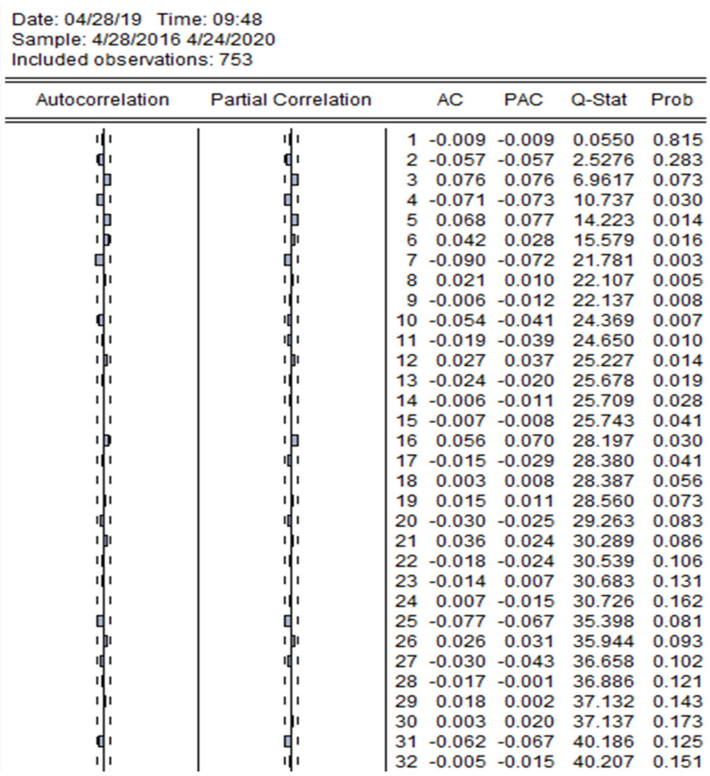


Figure 3. Graphical representation of the JNJ stock price index after differencing

The series "DP" of the JNJ index becomes stationary after the first difference, as shown in Figure 3 of the line graph and then in the correlogram, respectively.

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



Date: 04/28/19 Time: 09:48
Sample: 4/28/2016 4/24/2020
Included observations: 753

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.009	-0.009	0.0550	0.815
		2	-0.057	-0.057	2.5276	0.283
		3	0.076	0.076	6.9617	0.073
		4	-0.071	-0.073	10.737	0.030
		5	0.068	0.077	14.223	0.014
		6	0.042	0.028	15.579	0.016
		7	-0.090	-0.072	21.781	0.003
		8	0.021	0.010	22.107	0.005
		9	-0.006	-0.012	22.137	0.008
		10	-0.054	-0.041	24.369	0.007
		11	-0.019	-0.039	24.650	0.010
		12	0.027	0.037	25.227	0.014
		13	-0.024	-0.020	25.678	0.019
		14	-0.006	-0.011	25.709	0.028
		15	-0.007	-0.008	25.743	0.041
		16	0.056	0.070	28.197	0.030
		17	-0.015	-0.029	28.380	0.041
		18	0.003	0.008	28.387	0.056
		19	0.015	0.011	28.560	0.073
		20	-0.030	-0.025	29.263	0.083
		21	0.036	0.024	30.289	0.086
		22	-0.018	-0.024	30.539	0.106
		23	-0.014	0.007	30.683	0.131
		24	0.007	-0.015	30.726	0.162
		25	-0.077	-0.067	35.398	0.081
		26	0.026	0.031	35.944	0.093
		27	-0.030	-0.043	36.658	0.102
		28	-0.017	-0.001	36.886	0.121
		29	0.018	0.002	37.132	0.143
		30	0.003	0.020	37.137	0.173
		31	-0.062	-0.067	40.186	0.125
		32	-0.005	-0.015	40.207	0.151

Figure 4. The correlogram of JNJ stock price index after first differencing

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

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

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

Null Hypothesis: D(P) has a unit root
Exogenous: None
Lag Length: 0 (Automatic - based on SIC, maxlag=19)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-27.61292	0.0000
Test critical values:	1% level	-2.568041	
	5% level	-1.941245	
	10% level	-1.616416	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(P,2)
Method: Least Squares
Date: 04/27/19 Time: 21:15
Sample (adjusted): 5/02/2016 4/26/2019
Included observations: 752 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(P(-1))	-1.007680	0.036493	-27.61292	0.0000

R-squared	0.503790	Mean dependent var	1.18E-05
Adjusted R-squared	0.503790	S.D. dependent var	0.014374
S.E. of regression	0.010126	Akaike info criterion	-6.346171
Sum squared resid	0.076998	Schwarz criterion	-6.340023
Log likelihood	2387.160	Hannan-Quinn criter.	-6.343802
Durbin-Watson stat	1.999494		

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: D(P)
Method: ARMA Maximum Likelihood (OPG - BHHH)
Date: 04/27/19 Time: 21:25
Sample: 4/29/2016 4/26/2019
Included observations: 753
Convergence achieved after 32 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.000299	0.000511	0.584433	0.5591
AR(3)	0.338323	0.255723	1.323004	0.1862
MA(3)	-0.255770	0.269992	-0.947325	0.3438
MA(5)	0.077903	0.035174	2.214809	0.0271
SIGMASQ	0.000101	1.84E-06	54.87640	0.0000

R-squared	0.012750	Mean dependent var	0.000294
Adjusted R-squared	0.007470	S.D. dependent var	0.010116
S.E. of regression	0.010078	Akaike info criterion	-6.350257
Sum squared resid	0.075969	Schwarz criterion	-6.319552
Log likelihood	2395.872	Hannan-Quinn criter.	-6.338428
F-statistic	2.414983	Durbin-Watson stat	1.984069
Prob(F-statistic)	0.047518		

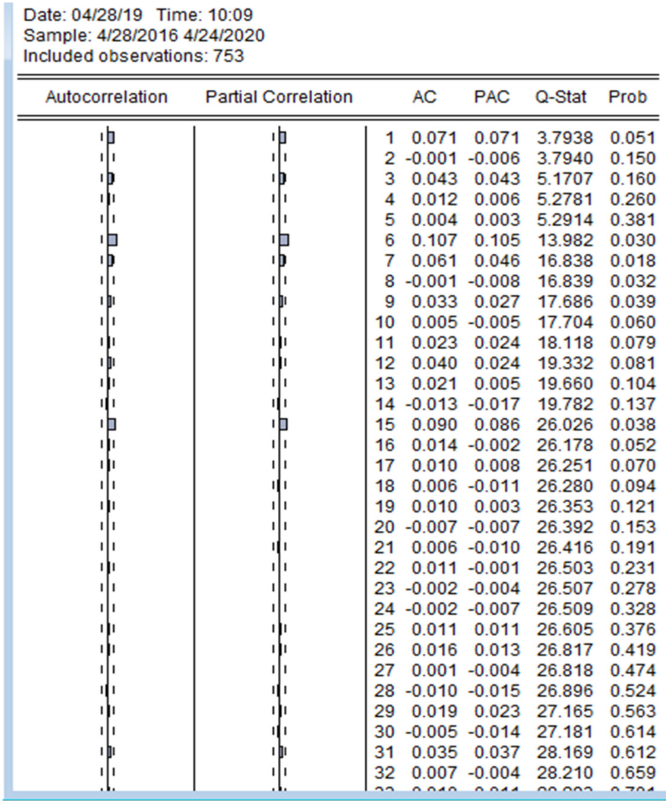
Inverted AR Roots	.70	-.35+.60i	-.35-.60i	
Inverted MA Roots	.53-.21i	.53+.21i	-.29+.66i	-.29-.66i
	-.47			

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
Q-statistic probabilities adjusted for 3 ARMA terms

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.008	0.008	0.0454	
		2	-0.054	-0.054	2.2388	
		3	-0.006	-0.005	2.2627	
		4	-0.060	-0.063	5.0011	0.025
		5	-0.001	-0.000	5.0014	0.082
		6	0.018	0.011	5.2524	0.154
		7	-0.079	-0.080	10.010	0.040
		8	0.000	-0.001	10.010	0.075
		9	-0.013	-0.023	10.147	0.119
		10	-0.044	-0.043	11.607	0.114
		11	-0.032	-0.044	12.400	0.134
		12	0.035	0.030	13.329	0.148
		13	-0.023	-0.029	13.737	0.185
		14	-0.005	-0.014	13.755	0.247
		15	-0.004	-0.011	13.764	0.316
		16	0.061	0.063	16.634	0.217
		17	-0.013	-0.025	16.765	0.269
		18	0.003	0.002	16.773	0.333
		19	0.016	0.019	16.970	0.388
		20	-0.023	-0.023	17.390	0.428
		21	0.030	0.028	18.086	0.450
		22	-0.010	-0.015	18.157	0.512
		23	-0.012	0.004	18.276	0.569
		24	0.003	-0.008	18.284	0.631
		25	-0.073	-0.069	22.437	0.434
		26	0.030	0.038	23.131	0.453
		27	-0.033	-0.045	23.981	0.463
		28	-0.004	-0.001	23.991	0.520
		29	0.019	0.010	24.263	0.561
		30	0.005	0.007	24.286	0.614
		31	-0.057	-0.061	26.843	0.527

Figure 7. Correlogram of residuals of the JNJ stock index



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

	AC	PAC	Q-Stat	Prob
1	0.071	0.071	3.7938	0.051
2	-0.001	-0.006	3.7940	0.150
3	0.043	0.043	5.1707	0.160
4	0.012	0.006	5.2781	0.260
5	0.004	0.003	5.2914	0.381
6	0.107	0.105	13.982	0.030
7	0.061	0.046	16.838	0.018
8	-0.001	-0.008	16.839	0.032
9	0.033	0.027	17.686	0.039
10	0.005	-0.005	17.704	0.060
11	0.023	0.024	18.118	0.079
12	0.040	0.024	19.332	0.081
13	0.021	0.005	19.660	0.104
14	-0.013	-0.017	19.782	0.137
15	0.090	0.086	26.026	0.038
16	0.014	-0.002	26.178	0.052
17	0.010	0.008	26.251	0.070
18	0.006	-0.011	26.280	0.094
19	0.010	0.003	26.353	0.121
20	-0.007	-0.007	26.392	0.153
21	0.006	-0.010	26.416	0.191
22	0.011	-0.001	26.503	0.231
23	-0.002	-0.004	26.507	0.278
24	-0.002	-0.007	26.509	0.328
25	0.011	0.011	26.605	0.376
26	0.016	0.013	26.817	0.419
27	0.001	-0.004	26.818	0.474
28	-0.010	-0.015	26.896	0.524
29	0.019	0.023	27.165	0.563
30	-0.005	-0.014	27.181	0.614
31	0.035	0.037	28.169	0.612
32	0.007	-0.004	28.210	0.659

Figure 8. Correlogram of square residuals of the JNJ stock index

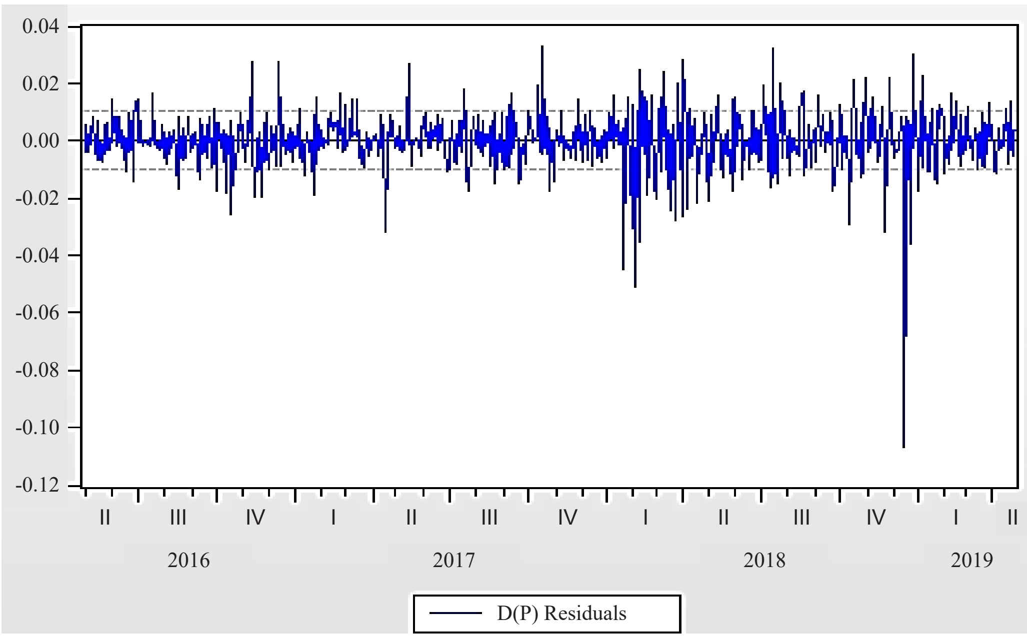


Figure 9. Graph of d(p) residuals of the JNJ stock index

3.1.1 *Forcasting of ARIMA AR (3), MA (3), MA (5)*

The optimal model selected can be expressed as follows in forecasting form:

$$P_t = a + \beta_0 P_{t-3} + \beta_1 E_{t-3} + \beta_2 E_{t-5} + e_t \tag{1}$$

Where $E_t = P_t - P_t$ (i.e., the difference between the actual value of the series and the forecast value).

Heteroskedasticity Test: ARCH

F-statistic	3.782966	Prob. F(1,750)	0.0522
Obs*R-squared	3.774018	Prob. Chi-Square(1)	0.0521

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
Date: 04/28/19 Time: 10:10
Sample (adjusted): 5/02/2016 4/26/2019
Included observations: 752 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	9.38E-05	1.70E-05	5.526335	0.0000
RESID^2(-1)	0.070842	0.036423	1.944985	0.0522

R-squared	0.005019	Mean dependent var	0.000101
Adjusted R-squared	0.003692	S.D. dependent var	0.000455
S.E. of regression	0.000455	Akaike info criterion	-12.55150
Sum squared resid	0.000155	Schwarz criterion	-12.53921
Log likelihood	4721.365	Hannan-Quinn criter.	-12.54677
F-statistic	3.782966	Durbin-Watson stat	1.999212
Prob(F-statistic)	0.052150		

Figure 10. Test of residual diagnostic

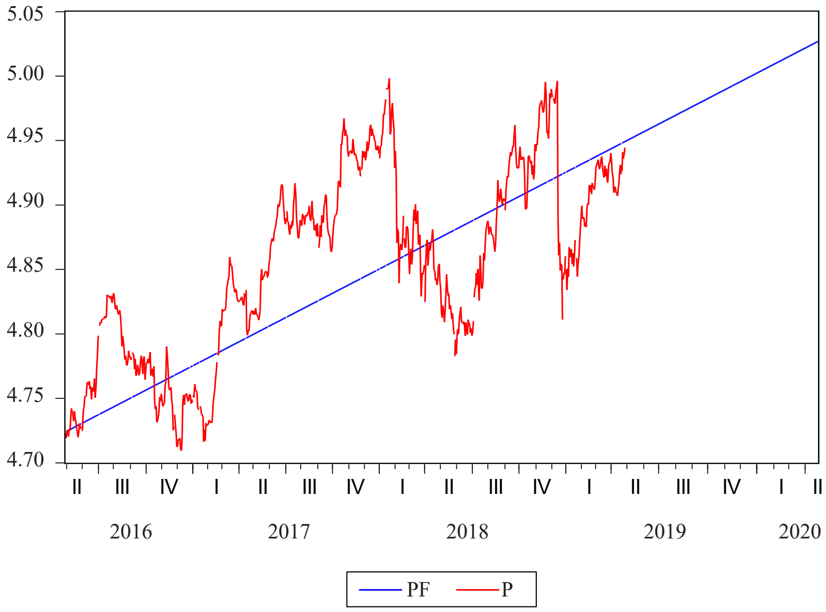


Figure 11. graph of P (actual sample) and PF (estimated sample)

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 2. Sample of empirical results of ARIMA AR (3) MA (3) MA (5) of JNJ stock index

Sample Period	P	PF
4/01/2019	4.934330	4.943743
4/02/2019	4.925150	4.944042
4/03/2019	4.921221	4.944341
4/04/2019	4.909488	4.944640
4/05/2019	4.913977	4.944939
4/08/2019	4.913684	4.945238
4/09/2019	4.909488	4.945537
4/010/2019	4.909652	4.945836
4/11/2019	4.906829	4.946135
4/12/2019	4.912508	4.946434
4/15/2019	43916471	4.946733
4/16/2019	4.927399	4.947031
4/17/2019	4.931015	4.947330
4/18/2019	4.923769	4.947629
4/22/2019	4.926021	4.947928
4/23/2019	4.940928	4.948227
4/24/2019	4.935912	4.948526
4/25/2019	4.939139	4.948825
4/26/2019	4.944424	4.949124
4/29/2019	NA	4.949423
4/30/2019	NA	4.949722
5/01/2019	NA	4.150021

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.

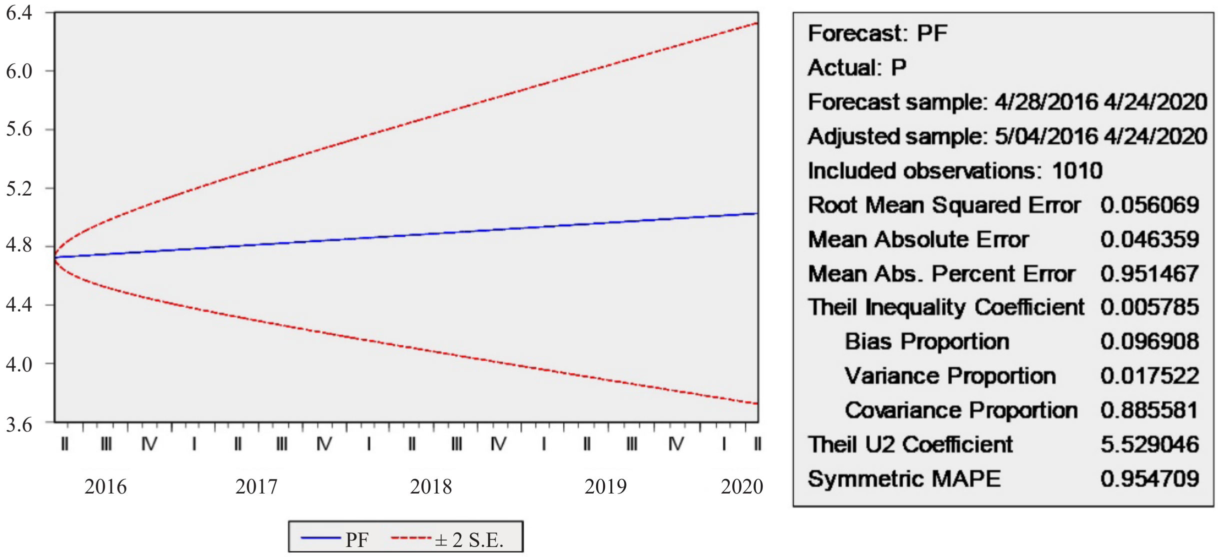


Figure 12. Graph of P dynamic (closing price actual) and PF (forecast price) of the JNJ stock index

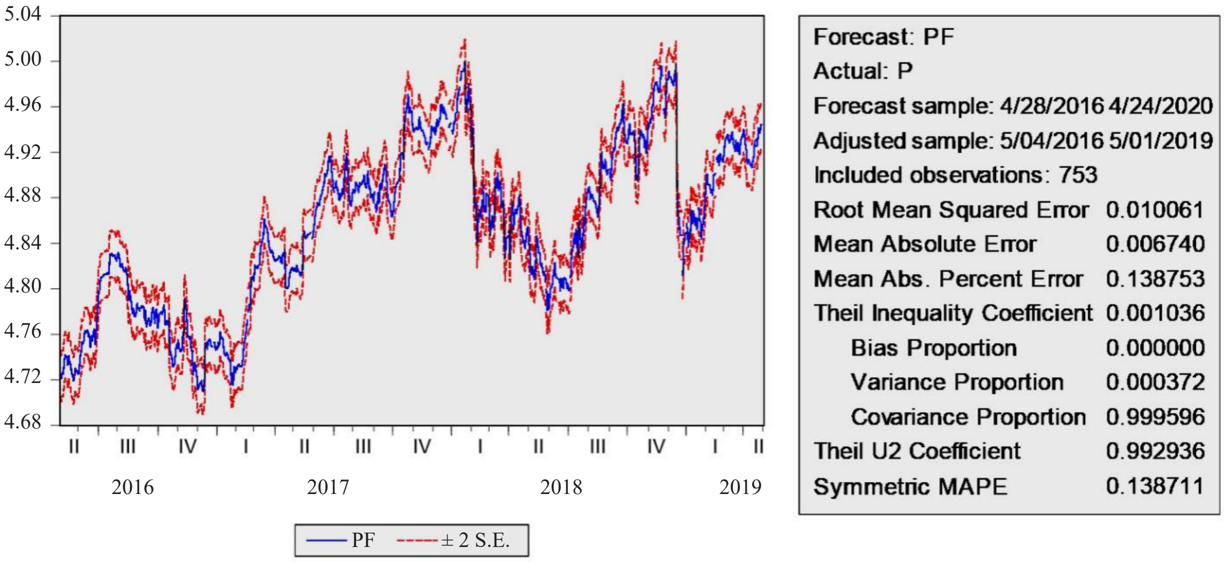
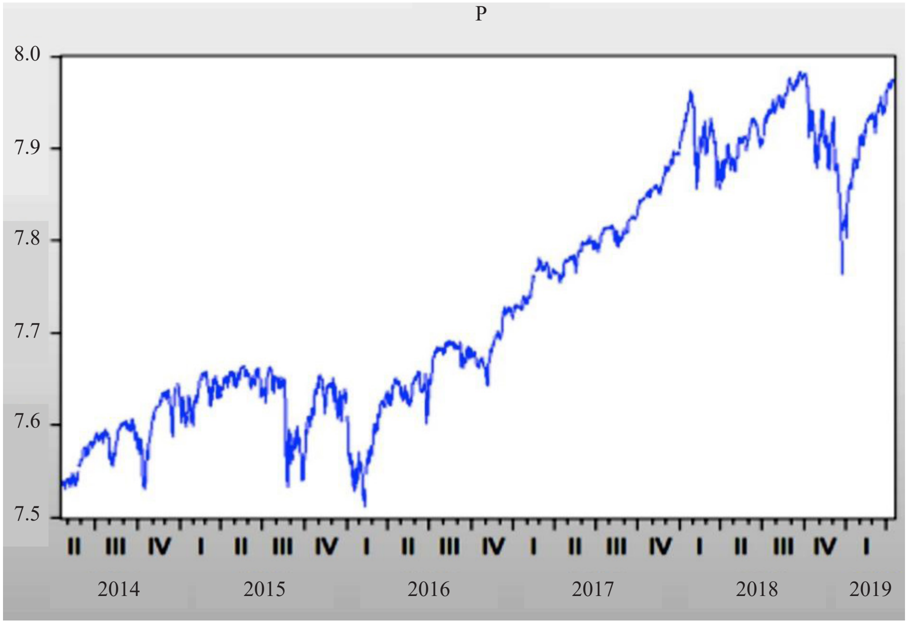


Figure 13. Graph of P static (closing price actual) and PF (forecast price) of the JNJ stock index

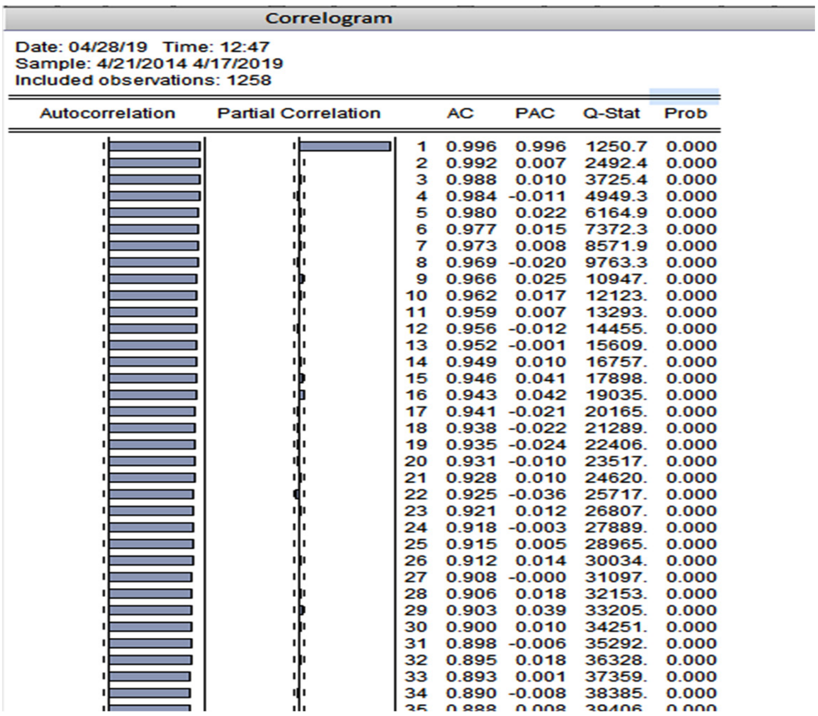
3.2 *ARIMA (p, d, q) model for S & P 500 index*

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.



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



Correlogram

Date: 04/28/19 Time: 12:47
Sample: 4/21/2014 4/17/2019
Included observations: 1258

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.996	0.996	1250.7	0.000
		2	0.992	0.007	2492.4	0.000
		3	0.988	0.010	3725.4	0.000
		4	0.984	-0.011	4949.3	0.000
		5	0.980	0.022	6164.9	0.000
		6	0.977	0.015	7372.3	0.000
		7	0.973	0.008	8571.9	0.000
		8	0.969	-0.020	9763.3	0.000
		9	0.966	0.025	10947.	0.000
		10	0.962	0.017	12123.	0.000
		11	0.959	0.007	13293.	0.000
		12	0.956	-0.012	14455.	0.000
		13	0.952	-0.001	15609.	0.000
		14	0.949	0.010	16757.	0.000
		15	0.946	0.041	17898.	0.000
		16	0.943	0.042	19035.	0.000
		17	0.941	-0.021	20165.	0.000
		18	0.938	-0.022	21289.	0.000
		19	0.935	-0.024	22406.	0.000
		20	0.931	-0.010	23517.	0.000
		21	0.928	0.010	24620.	0.000
		22	0.925	-0.036	25717.	0.000
		23	0.921	0.012	26807.	0.000
		24	0.918	-0.003	27889.	0.000
		25	0.915	0.005	28965.	0.000
		26	0.912	0.014	30034.	0.000
		27	0.908	-0.000	31097.	0.000
		28	0.906	0.018	32153.	0.000
		29	0.903	0.039	33205.	0.000
		30	0.900	0.010	34251.	0.000
		31	0.898	-0.006	35292.	0.000
		32	0.895	0.018	36328.	0.000
		33	0.893	0.001	37359.	0.000
		34	0.890	-0.008	38385.	0.000
		35	0.888	0.008	39406.	0.000

Figure 15. The correlogram of S & P 500 index

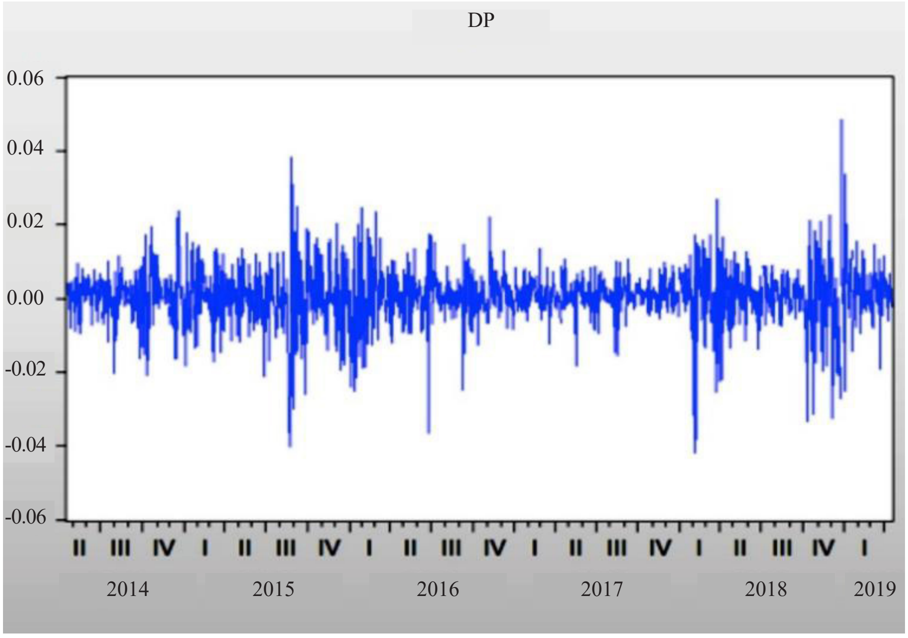


Figure 16. Graphical representation of the S & P 500 index after differencing

Date: 04/28/19 Time: 12:51
Sample: 4/21/2014 4/17/2019
Included observations: 1257

Correlogram DP

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.017	-0.017	0.3731	0.541
		2	-0.030	-0.030	1.4961	0.473
		3	0.020	0.019	1.9913	0.574
		4	-0.053	-0.053	5.5203	0.238
		5	-0.029	-0.030	6.6036	0.252
		6	-0.001	-0.006	6.6046	0.359
		7	0.027	0.027	7.5011	0.379
		8	-0.055	-0.056	11.316	0.184
		9	-0.029	-0.033	12.411	0.191
		10	-0.006	-0.013	12.465	0.255
		11	0.004	0.006	12.481	0.329
		12	0.010	0.006	12.610	0.398
		13	0.004	-0.001	12.633	0.477
		14	-0.064	-0.069	17.873	0.213
		15	-0.061	-0.062	22.579	0.093
		16	0.031	0.024	23.778	0.094
		17	0.052	0.051	27.270	0.054
		18	0.026	0.023	28.126	0.060
		19	-0.020	-0.030	28.654	0.072
		20	0.013	0.010	28.856	0.091
		21	0.043	0.054	31.270	0.069
		22	-0.016	-0.009	31.581	0.085
		23	0.022	0.011	32.204	0.096
		24	0.013	0.005	32.408	0.117
		25	-0.028	-0.016	33.432	0.121
		26	0.006	0.017	33.484	0.149
		27	-0.024	-0.023	34.222	0.160
		28	-0.032	-0.038	35.511	0.155
		29	-0.021	-0.030	36.067	0.172
		30	-0.011	-0.013	36.234	0.201
		31	-0.063	-0.054	41.296	0.102
		32	0.004	0.008	41.319	0.125
		33	0.017	0.003	41.694	0.143
		34	0.003	-0.002	41.707	0.171
		35	-0.016	-0.018	42.022	0.193

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.

Augmented Dickey-Fuller Unit Root Test on D(P)

Null Hypothesis: D(P) has a unit root
Exogenous: None
Lag Length: 0 (Automatic - based on SIC, maxlag=22)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-35.97993	0.0000
Test critical values:	1% level	-2.566804	
	5% level	-1.941076	
	10% level	-1.616530	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(P,2)
Method: Least Squares
Date: 04/28/19 Time: 12:50
Sample (adjusted): 4/23/2014 4/17/2019
Included observations: 1256 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(P(-1))	-1.015449	0.028223	-35.97993	0.0000

R-squared	0.507757	Mean dependent var	-5.06E-06
Adjusted R-squared	0.507757	S.D. dependent var	0.011923
S.E. of regression	0.008365	Akaike info criterion	-6.728760
Sum squared resid	0.087813	Schwarz criterion	-6.724671
Log likelihood	4226.661	Hannan-Quinn criter.	-6.727223
Durbin-Watson stat	2.000563		

Figure 18. ADF unit root test for DP of S & P 500 index

Table 3. Statistical results of different Arima parameters for S & P index

Differenced adj Closing price	ARIMA (p, d, q) (4, 1, 4)	ARIMA (p, d, q) (4, 1, 8)	ARIMA (p, d, q) (8, 1, 4)	ARIMA (p, d, q) (8, 1, 8)
Sigma sq Volatility	6.95E-05	6.94E-05	6.93E-05	6.95E-05
Adj R sq	0.002099	0.003531	0.003717	0.001076
Akaike info criterion	-6.730569	-6.731995	-6.732179	-6.729538
Schwarz criterion	-6.714224	-6.715650	-6.715834	-6.713193

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.

Dependent Variable: D(P)
Method: ARMA Maximum Likelihood (OPG - BHHH)
Date: 04/28/19 Time: 12:56
Sample: 4/22/2014 4/17/2019
Included observations: 1257
Convergence achieved after 19 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.000348	0.000226	1.538460	0.1242
AR(8)	-0.054223	0.022320	-2.429378	0.0153
MA(4)	-0.055402	0.020263	-2.734128	0.0063
SIGMASQ	6.93E-05	1.67E-06	41.64579	0.0000

R-squared	0.006097	Mean dependent var	0.000348
Adjusted R-squared	0.003717	S.D. dependent var	0.008356
S.E. of regression	0.008341	Akaike info criterion	-6.732179
Sum squared resid	0.087164	Schwarz criterion	-6.715834
Log likelihood	4235.175	Hannan-Quinn criter.	-6.726036
F-statistic	2.562019	Durbin-Watson stat	2.035984
Prob(F-statistic)	0.053452		

Inverted AR Roots	.64-.27i	.64+.27i	.27+.64i	.27-.64i
	-.27-.64i	-.27+.64i	-.64+.27i	-.64-.27i
Inverted MA Roots	.49	-.00+.49i	-.00-.49i	-.49

Figure 19. ARIMA (8, 1, 4) estimation output with D(P) of S & P 500 index

Date: 04/28/19 Time: 13:02
Sample: 4/21/2014 4/17/2019
Included observations: 1257
Q-statistic probabilities adjusted for 2 ARMA terms

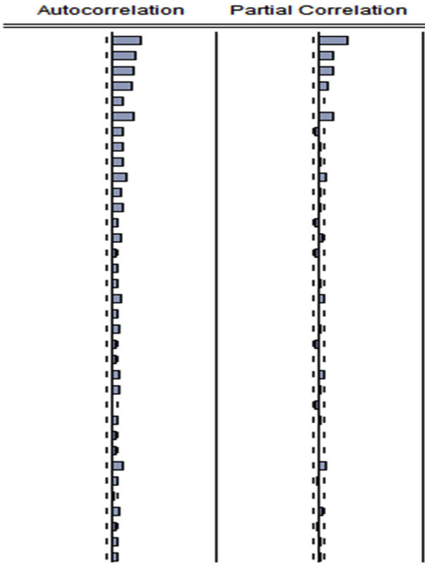
Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.018	-0.018	0.4129	
		2	-0.033	-0.033	1.7621	
		3	0.019	0.018	2.2108	0.137
		4	-0.000	-0.001	2.2109	0.331
		5	-0.031	-0.030	3.4024	0.334
		6	-0.008	-0.010	3.4930	0.479
		7	0.024	0.022	4.2052	0.520
		8	0.001	0.003	4.2073	0.649
		9	-0.029	-0.027	5.2745	0.627
		10	-0.011	-0.013	5.4223	0.712
		11	0.002	-0.001	5.4257	0.796
		12	0.010	0.011	5.5444	0.852
		13	0.006	0.008	5.5953	0.899
		14	-0.065	-0.066	10.947	0.533
		15	-0.060	-0.064	15.495	0.277
		16	0.030	0.025	16.634	0.276
		17	0.052	0.054	20.112	0.168
		18	0.021	0.027	20.702	0.190
		19	-0.024	-0.027	21.445	0.207
		20	0.014	0.006	21.689	0.246
		21	0.044	0.048	24.221	0.188
		22	-0.018	-0.007	24.650	0.215
		23	0.013	0.012	24.854	0.254
		24	0.014	0.002	25.095	0.293
		25	-0.023	-0.023	25.794	0.311
		26	0.006	0.015	25.845	0.361
		27	-0.029	-0.026	26.908	0.360
		28	-0.029	-0.037	27.996	0.359
		29	-0.022	-0.035	28.602	0.380
		30	-0.011	-0.013	28.768	0.424
		31	-0.062	-0.053	33.789	0.247
		32	0.007	0.013	33.848	0.287
		33	0.012	0.003	34.049	0.323
		34	0.003	-0.002	34.061	0.369

Figure 20. 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.

Date: 04/28/19 Time: 13:02
Sample: 4/21/2014 4/17/2019
Included observations: 1257



	AC	PAC	Q-Stat	Prob
1	0.292	0.292	107.36	0.000
2	0.227	0.155	172.15	0.000
3	0.220	0.133	233.32	0.000
4	0.192	0.086	279.83	0.000
5	0.114	-0.003	296.23	0.000
6	0.217	0.146	355.91	0.000
7	0.100	-0.032	368.46	0.000
8	0.100	0.015	381.04	0.000
9	0.103	0.022	394.56	0.000
10	0.140	0.068	419.60	0.000
11	0.097	0.020	431.46	0.000
12	0.104	0.010	445.15	0.000
13	0.054	-0.025	448.81	0.000
14	0.090	0.035	459.18	0.000
15	0.030	-0.041	460.29	0.000
16	0.047	-0.004	463.10	0.000
17	0.054	0.020	466.80	0.000
18	0.094	0.057	478.02	0.000
19	0.053	0.007	481.60	0.000
20	0.077	0.015	489.10	0.000
21	0.025	-0.030	489.90	0.000
22	0.041	0.002	492.04	0.000
23	0.073	0.047	498.83	0.000
24	0.069	0.010	504.93	0.000
25	0.007	-0.037	505.00	0.000
26	0.049	0.017	508.06	0.000
27	0.028	-0.001	509.07	0.000
28	0.035	0.003	510.64	0.000
29	0.102	0.079	523.98	0.000
30	0.054	-0.017	527.67	0.000
31	0.023	-0.008	528.36	0.000
32	0.080	0.041	536.72	0.000
33	0.036	-0.018	538.43	0.000
34	0.044	0.013	540.96	0.000
35	0.063	0.014	546.04	0.000

Figure 21. Correlogram of residuals of the S & P 500 index

Heteroskedasticity Test: ARCH

F-statistic	116.8224	Prob. F(1,1254)	0.0000
Obs*R-squared	107.0371	Prob. Chi-Square(1)	0.0000

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
Date: 04/28/19 Time: 13:03
Sample (adjusted): 4/23/2014 4/17/2019
Included observations: 1256 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	4.91E-05	4.93E-06	9.973366	0.0000
RESID^2(-1)	0.291930	0.027009	10.80844	0.0000

R-squared	0.085221	Mean dependent var	6.94E-05
Adjusted R-squared	0.084491	S.D. dependent var	0.000169
S.E. of regression	0.000161	Akaike info criterion	-14.62320
Sum squared resid	3.27E-05	Schwarz criterion	-14.61502
Log likelihood	9185.366	Hannan-Quinn criter.	-14.62012
F-statistic	116.8224	Durbin-Watson stat	2.090349
Prob(F-statistic)	0.000000		

Figure 22. Test of heteroskedasticity of ARIMA (8, 1, 4) of S & P 500 index

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

Alwadi, R. (2015). Forecasting short term financial data. *European Scientific Journal, 11*(25).

Al Wadia, M. T. I. S., & Ismail, M. T. (2011). Selecting wavelet transforms model in forecasting financial time series data based on ARIMA model. *Applied Mathematical Sciences, 5*(7), 315-326.

Al Wadi, S., Ismail, M. T., Altaher, A. M., & Karim, S. A. A. (2010, December). Forecasting volatility data based on Wavelet transforms and ARIMA model. In *2010 International Conference on Science and Social Research (CSSR 2010)* (pp. 86-90). IEEE.

Al Wadi, S., Hamarsheh, A., & Alwadi, H. (2013). Maximum overlapping discrete wavelet transform in forecasting banking sector. *Applied Mathematical Sciences, 7*(80), 3995-4002.

Asteriou, D., & Hall, S. G. (2007). Applied Econometrics: a modern approach, revised edition. *Hampshire: Palgrave Macmillan, 46*(2), 117-155.

Atsalakis, G. S., & Kimon, P. V. (2009). Forecasting stock market short-term trends using a neuro-fuzzy methodology. *Expert Systems with Applications, 36*(7), 10696-10707.

Box, G. E. (1970). *GM Jenkins Time Series Analysis: Forecasting and Control*. San Francisco, Holdan-Day.

Box, G. P., & Jenkins, G. M. (1976). *Time Series Analysis: Forecasting and Control*. Holden-day Inc., San Francisco, CA.

Bozarth, C. B., & Handfield, R. B. (2016). *Introduction to Operations and Supply Chain Management, eBook*. Pearson Higher Ed.

Fattah, J., Ezzine, L., Aman, Z., El Moussami, H., & Lachhab, A. (2018). Forecasting of demand using ARIMA model. *International Journal of Engineering Business Management, 10*, 1847979018808673.

Gujarati, D. N., Porter, D. C., & Gunasekar, S. (2012). *Basic Econometrics*. Tata mcgraw-hill education.

Ho, S. L., & Xie, M. (1998). The use of ARIMA models for reliability forecasting and analysis. *Computers & Industrial Engineering, 35*(1-2), 213-216.

Jaber, J. J., Ismail, N., Wadi, S. A., & Saleh, M. H. (2017). Forecasting of volatility risk for Jordanian banking sector. *Far East Journal of Mathematical Sciences, 101*(7), 1491-1507.

Johnson & Johnson (JNJ) NYSE-NYSE Closing Price. (2019, April 26). Currency in USD. Yahoo Finance. https://finance.yahoo.com/quote/JNJ/history?p=JNJ

Mahir, R. A., & Al-Khazaleh, A. M. H. (2008). Estimation of missing data by using the filtering process in a time series modeling. https://doi.org/10.48550/arXiv.0811.0659

S & P 500 (^GSPC) SNP-SNP Real Time Price. (2019, April 17). Currency in USD.

Yahoo Finance (2019, April 17). https://finance.yahoo.com/quote/%5EGSPC/history?p=%5EGSPC

Tsay, R. S. (2000). Time series and forecasting: Brief history and future research. *Journal of the American Statistical Association, 95*(450), 638-643.

Wooldridge, J. M. (2009). *Introductory Economics: A Modern Approach*. Mason, Cengage-Learning.