Enhanced Knowledge in Sciences and Technology Vol. 3 No. 1 (2023) 001-010

EKST

Homepage: http://publisher.uthm.edu.my/periodicals/index.php/ekst
e-ISSN : 2773-6385

The FGV Stock Price Forecasting Using Holt's Linear Trend and ARIMA Model

Rabia'tul Adawiyah Sua'aif[1] and Norhaidah Mohd Asrah[2*]

[1, 2] Faculty of Applied Sciences and Technology,
Universiti Tun Hussein Onn Malaysia (Pagoh Campus),
84600 Pagoh, Muar, Johor, MALAYSIA.

*Corresponding Author Designation

DOI: https://doi.org/10.30880/ekst.2023.03.01.001
Received 15 September 2022; Accepted 18 December 2022; Available online 3 August 2023

Abstract: This study analyses the stock price performance of Felda Global Venture Holding Berhad (FGV), the largest palm oil producer. Due to the non-consistency and faced declining stock price, the prediction of the stock price represented in this study leads to the determination of an appropriate forecast model. In this study, the daily stock price of FGV is examined using two forecasting techniques, their forecasting accuracy is compared, and the FGV stock price is forecasted using the most accurate technique. Two forecasting methods are used in this study known as Holt's linear trend and the ARIMA model. Datasets consist of 1919 data of observations of FGV daily stock price from 2nd January 2013 until 30th October 2020. The data were split into two portions where 860 was the training set and 59 was the testing set. The result revealed that ARIMA (0, 1, 2) is the best forecasting model for FGV stock price prediction with the lowest error in the accuracy measurements compared to Holt's linear trend model. This study is recommended for applying advanced forecasting techniques to the issues that arise nowadays for further analysis due to evaluating the issues affecting the stock price behaviour.

Keywords: FGV Holding Berhad, Stock Price Forecasting, ARIMA, Holt's Linear Trend

1. Introduction

Time series analysis is a statistical technique that examines time series data that are gathered at regular intervals of time and may be used to forecast future occurrences. The act of prediction is called forecasting while the prediction is known as forecast. The data on stock price is one of the significant data that impact investors and businesses in the future. Every stock price has a unique behaviour, and in order to predict that behaviour in a way that is useful to the market and will contribute to the future, there are a few concepts that must be understood [1]. Hence, Felda Global Venture Holding Berhad (FGV), the largest firm with palm oil as its primary product, is chosen to a study that examines stock price performance. As one of Malaysia's leading plantation firms, FGV stock price forecasting is crucial for presenting the business's performance on the market. The FGV is producing crude palm oil and was

*Corresponding author: norhaida@uthm.edu.my

a private limited company before operating as the commercial arm of the Felda Land Development Authority (FELDA) in 2007 and becoming the main market of Bursa Malaysia Securities Berhad as Felda Global Venture Holding Berhad on 28th June 2012 [2].

The goal of the analysis of FGV stock price is to comprehend and examine stock price changes because of FGV's lack of consistency and stock price decline over a period of seven years. Fundamentally, every company needs to consider its performance, trend, and stock market behaviour when making stock price predictions [3]. Thus, the analysis and prediction in this study are as a trial to discover the FGV stock price value in the future. In the spite of that, the appraisement of the FGV stock price prediction is motionless if an accurate forecasting technique is not proposed well due to deciding to apply the right forecasting techniques is lead to the essential supporter of effective planning to gain profits in the future. Also, the forecasting model is presented to preview the time series movement, and the technical analysis method is one of the bases for predicting future stock prices [4]. Therefore, the forecast techniques for FGV stock price prediction are presented in this study which leading the determination of an appropriate forecast model.

Forecasting became an important activity in the business field to analyse the prediction of products and services requested in the future [5]. Forecasting was helpful in various types of organizations because the applications will spread many benefits and convenience [6]. ARIMA is a consolidation of the Autoregressive model (AR) and Moving Average model (MA). Meanwhile, the letter 'I' between AR and MA stood for Integrated [7]. Exponential smoothing is also widely used in time series analysis and forecasting of the future due to it is an easy and simple calculation [8][9]. As a result, the purpose of this study is to assess the daily stock price of FGV using the Box-Jenkins model and Holt's linear trend, an exponential smoothing technique. Second, using mean absolute error (MAE), mean absolute percentage error (MAPE), mean square error (MSE), and root mean square error (RMSE), compare the predicting accuracy of various methodologies. Lastly, to predict the price of the FGV stock using the most effective forecasting methodology.

The data used in this study was secondary data obtained from Investing.com. The use of daily historical stock price data of FGV within seven years which are from 2013 until 2020. Forecasting methods for time series data, such as ARIMA and Holt's linear trend method. The goal of this study is to investigate and evaluate the performance, movement, and behaviour of the business's stock price in order to benefit FGV, the largest plantation company in Malaysia. Additionally, this research is focused on assisting FGV in forecasting stock prices and choosing the most effective model to represent stock prices in the future. The development of a country's economy will also benefit from learning about stock price prediction.

2. Methodology

2.1 Box-Cox transformation

This study performs a Box-Cox transformation to transform the dependent variable into normal form and logarithm, which helps stabilize the time series variance. The transformation of the Box-Jenkins method from non-stationary series to stationary series is important.

Table 1: The Box-Cox transformation table

Lambda value (λ)	Types of transformation	Transformed data
$\lambda = 0.50$	Square root transformation	$Y^{0.5} = \sqrt{Y}$

2.2 Stationary of time series

Over time, a fixed time series model will generate random patterns. Because the values of the time series at different times would be altered by the model's non-stationarity, it is crucial to ensure that the time series model is stationary [10]. Stationarity is also a major step in part identification and parameter selection of time series models [4]. As well to look at the time series plot of the data, the ACF plot is useful for identifying the non-stationary time series and another way to ensure the difference between the data is necessary or not by conducting the unit root test. There are two-unit root tests are conducted such as Augmented Dickey-Fuller (ADF) test and Kwiatkowski-Phillips-Schmidt-Shin (KPSS). These tests are described as statistical hypothesis tests of stationarity that are designed to examine whether differencing is needed.

The null hypothesis for the ADF test refers to the data being non-stationary, and the alternative hypothesis refers to the data being stationary. By using the 5% threshold, if the *p*-value is greater than 0.05 means that the differencing is required. For the KPSS test, the null hypothesis refers to the data being stationary while the alternative hypothesis refers to the data being non-stationary, if the *p*-value is smaller than 0.05 means that the data is not stationary, and differencing is required. Additionally, the stationarity of time series will be demonstrated via the differencing approach, which aids in stabilising the time series mean. Differentiation is an iterative procedure until a stationary time series is present. The abbreviation ARMA will be used if no differentiation is desired.

2.3 Autoregressive Integrated Moving Average (ARIMA)

The autoregressive model or usually called as AR model in ARIMA was proposed in this study and an autoregressive model of order *p* can be written as in *Eq.* 1*:*

$$y_t = c + \emptyset_1 y_{t-1} + \emptyset_2 y_{t-2} + \cdots + \emptyset_p y_{t-p} + e_t \qquad Eq.1$$

where y_t are the original series, $\emptyset$'s called are the autoregressive parameters to be estimated, *c* is the average of the changes between consecutive observations and e_t are refer to white noise. Moving average models use past forecast errors in regression-type models, in other words, MA models measure the adjustment of new forecasts to previous forecast errors [10]. Thus, the moving average model of order *q* can be written as in *Eq.* 2:

$$y_t = c + e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \cdots + \theta_q e_{t-q} \qquad Eq.\ 2$$

In ARIMA, model identification is required to choose the value of *p, d,* and *q* to refer to autoregressive order, differencing, and moving average order respectively in the ARIMA (*p, d, q*) model. The *p, d,* and *q* values will be estimated by understanding the autocorrelation function and partial autocorrelation function (ACF and PACF) plots. These two plots are adequate to determine an appropriate value for order *p* and *q* for AR and MA models respectively. The first model can be written as in *Eq.* 3:

$$y'_t = c + \emptyset_1 y'_{t-1} + \cdots + \emptyset_p y'_{t-p} + \emptyset_1 e_{t-1} + \cdots + \emptyset_q e_{t-q} + e_t \qquad Eq.\ 3$$

where y'_t is referred to differenced series and differencing may have been more than one time. While on the right side is known as the predictors, include both lagged values of y_t and lagged errors. Combining the model might be a complicated way to make it, however, it becomes easier to work with backshift notation as shown in *Eq. 4*:

$$1-\emptyset_1 B-\cdots-\emptyset_p B^p)(1-B)^d y_t = c+(1+\emptyset_1 B+\cdots+\emptyset_q B^q)e_t \qquad Eq.\ 4$$

The right-hand side refers to the MA(q) while the left-hand side indicates the AR(p) and d differencing respectively.

Utilizing statistical methods like the ACF and PACF, perform parameter estimates to identify potential suitable models [10]. Table 2 presents guidelines for identifying salient patterns using the ACF and PACF models.

Table 2: The model identification by using ACF and PACF patterns

Model	ACF patterns	PACF patterns
MA(q)	Cut off after lag q	Dies down or exponentially decaying
AR(p)	Dies down or exponentially decaying	Cut off after lag p

The least of Akaike's information criterion (AIC), corrected Akaike's information criterion (*AICc*), and Bayesian information criteria will be used to estimate and choose the best model (BIC). Analysis of the fitted model's residuals is another complimentary approach, and the residuals may be compared as illustrated in *Eq. 5*.

$$residual = actual\ value - predicted\ \text{value} \qquad Eq.\ 5$$

Since the model is significant, the parameter estimates are reasonably near to the real values, and the residuals should practically have white noise qualities. Additionally, autocorrelation plots will be used to look for additional structures with high correlation values. This model will be deemed suitable and used to make predictions because the residuals' autocorrelation (ACF) and partial autocorrelation (PACF) plots are small.

2.4 Holt's linear trend method

One of the exponential smoothing approaches that has been used is Holt's trend exponential smoothing, which is used when the time series is approximately growing or decreasing at a given rate [10].

$$y_t = \beta_0 + \beta_1 t + \varepsilon_t \qquad Eq.6$$

Eq. 6 shows that the parameter β_0 and β_1 are slowly changing over time. Therefore, regression will be used to the future value of y_t. There are two estimates which are a level estimate and the trend estimate as in *Eq. 7* and *8*.

$$l_t = \alpha y_t(1+\alpha)(l_{t-1}+b_{t-1}) \qquad Eq.\ 7$$
$$b_t = \beta(l_t - l_{t-1}) + (1-\beta)b_{t-1} \qquad Eq.\ 8$$

where l_t presented an estimate of the level of the series at the e t, and α is the smoothing constant for the level ($0\leq \alpha \leq 1$), b_t is denoted an estimate of the trend which is the slope of the series at time t, followed by β is indicate the smoothing constant for the trend ($0\leq\beta\leq 1$). Level estimate l_{t-1} which is the estimate of the level of the time series constructed in period $t-1$ that called the permanent component. Another one would be trend estimate, subscript which b_{t-1}which is the estimate of the growth rate of the time series constructed in the period $t-1$ that called a trend component.

$$\hat{y}_{+p}\ (T) = (l_T + pb_T), \qquad (p = 1,2,3,\ldots) \qquad Eq.9$$

For the point forecast made at time $for\ y_{T+p}$ as shown in *Eq. 9* where the *p*-step-ahead forecast is corresponding to the last value estimated level plus *p* times the last estimated trend value. Therefore, the forecast values for the next period are the linear function of *p*.

2.5 Forecast evaluation

The forecast will be evaluated by applying accuracy measurements such as mean squared error (MSE), root means squared error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE). The formulas for each measurement can be referred to *Eq. 10* until *Eq. 13*.

$$MSE = \text{the } \frac{1}{n}\sum_{i=1}^{n}(y - \hat{y}_I)^2 \qquad Eq.10$$

where n denoted the total number of observations, y_i is the actual value, $\hat{y}_i$ is the predicted value.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{N}} \qquad Eq.11$$

where n denoted the total number of observations, y_t is the actual value and $\hat{y}_t$ indicates the predicted value.

$$MAE = \frac{\sum_{i=1}^{n}|\hat{y}_i - y_i|}{n} \qquad Eq.12$$

where n denoted the total number of data points, y_t is the actual value, $\hat{y}_t$ is the prediction value.

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{y_i - \hat{y}_i}{y_i} \text{ x } 100 \qquad Eq.13$$

where n denoted the total number of data points, y is the actual value, y is the prediction value.

3. Results and Discussion

The time series plot for FGV is shown in Figure 1. The stock price for the first two years, from 2013 to 2020, dropped significantly from RM4.00 to RM2.00. From 2015 to 2018, the stock price then varied on average between RM1.00 and RM2.00. The stock price climbed and had an average price from the second half of 2019 to the third quarter of 2020, however it decreased and fell to less than RM1.00 in the interim. Due to the fact that the data looked to be somewhat decreasing over time, the stock price of FGV had a little curve in the time series data.

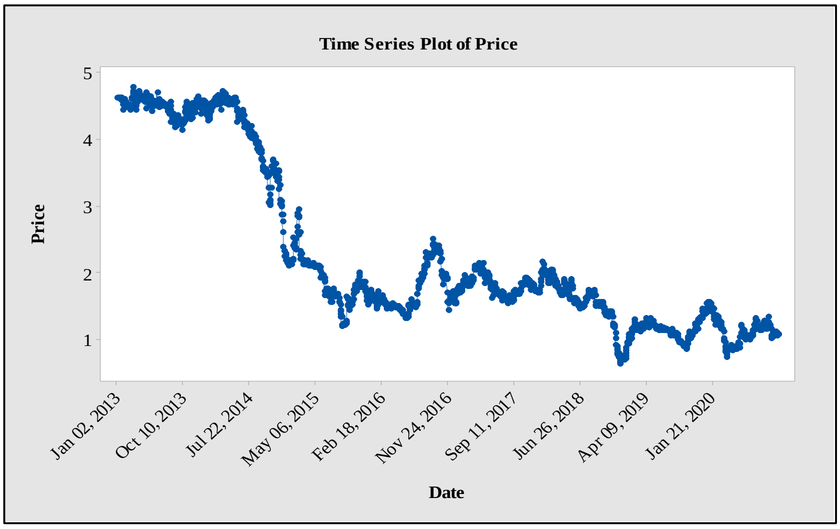
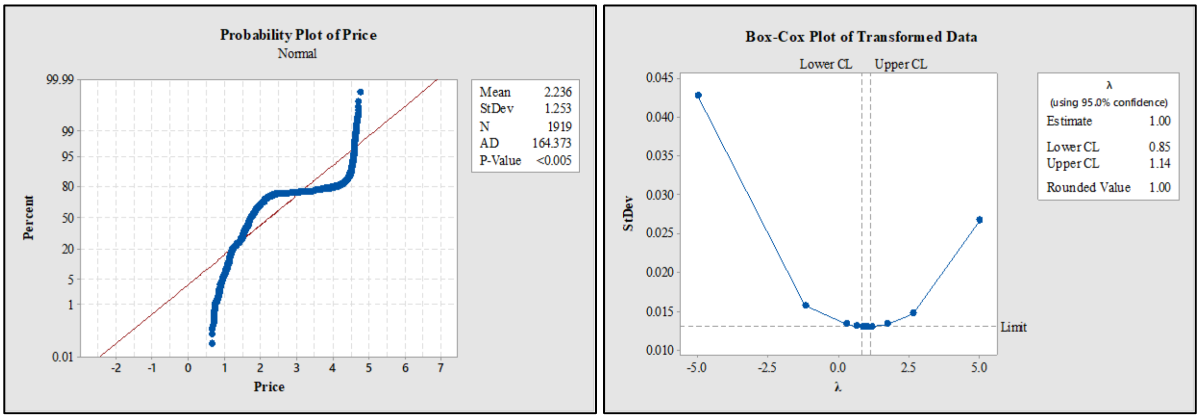


Figure 1: The time series plotting for FGV daily stock price from 2013 until 2020

The dataset was divided into training and testing data, due to how well a model fits toward the historical data. The training dataset was used for a preliminary study to determine an appropriate model and to determine how well a model chosen can performs on new data that was not used when fitting the model. While the testing dataset was used when choosing the model and measuring the model is likely to forecast on the new data. There was an 1860 training set used in this study and 59 number the testing set.



a b

Figure 2: (a) The normality distributed data plot and (b) the Box-Cox transformation

The data set is normal by comparing the *p*-value of the Anderson-Darling test as shown in Figure 2(a). Hypothesis testing is used in the decision-making of normality tests. As the null hypothesis, H_0 refer to the data do follow a normal distribution if the *p*-value is more than 0.05, while the alternative hypothesis, H_1 refer to data that do not follow a normal distribution if the *p*-value is less than 0.05 of the significance level. Based on Figure 2(a), that the *p*-value was less than 0.005, therefore, the decision is to reject the H_0 and can be concluded that, the data was not following the normal distribution. The Box-Cox transformation is then applied to square the variance using the square root transformation. The ideal lambda in Figure 2(b) is equal to 1, showing that the converted data is identical to the original data.

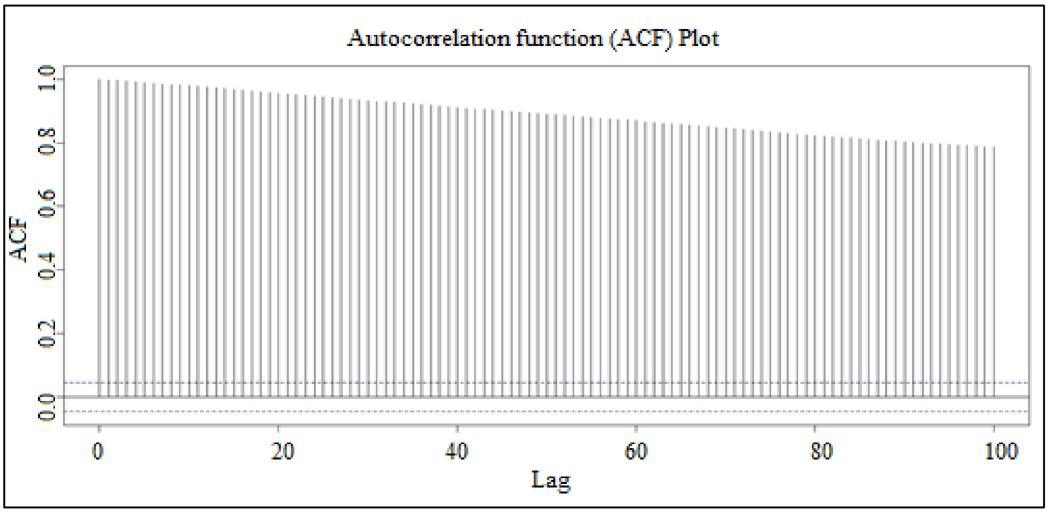


Figure 3: The ACF plot before data transformation

Figure 3 shows that the ACF plot before data transformation and ACF plot was a drop from lag 1 to lag 100 was quite slowly. Therefore, it can be concluded that the data was not stationary and differencing is needed. Besides, a unit root test was conducted to determine either differencing is required or not by applying the ADF test and KPSS test by using R Software as shown in Table 3 before and after conducting the differencing.

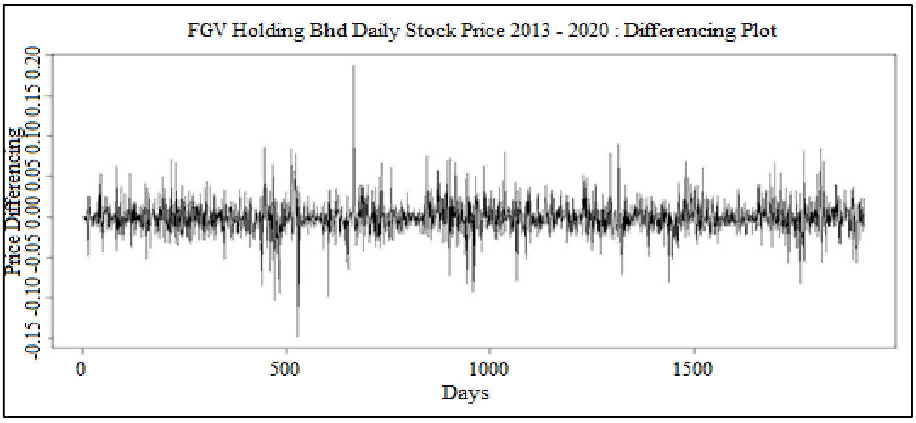


Figure 4: Differencing plot of FGV daily stock price from 2013 until 2020

Figure 4 demonstrates the differencing plot of the data and based on the time series stationary was appeared without the trend and the time series do not depend on the time. Therefore, after the first differencing conducted and the time series plot with stationary appeared, so that no differencing is required for the second time.

Table 3: The summary of the unit root test

Unit Root Test	*p*-value before differencing	Conclusion	*p*-value after differencing	Conclusion
Augmented Dickey-Fuller	0.6041	Not Stationary	0.0100	Stationary
Kwiatkowski-Phillips-Schmidt-Shin	0.010	Not Stationary	0.100	Stationary

Based on Table 3, before data differencing, the *p*-value of each test did not follow the null hypothesis and the data are not-stationary. Hence, differencing is required and, after conducted the differencing, the stationary had appeared the *p*-value of the ADF test is less than 0.05 while the *p*-value of the KPSS test is more than 0.05.

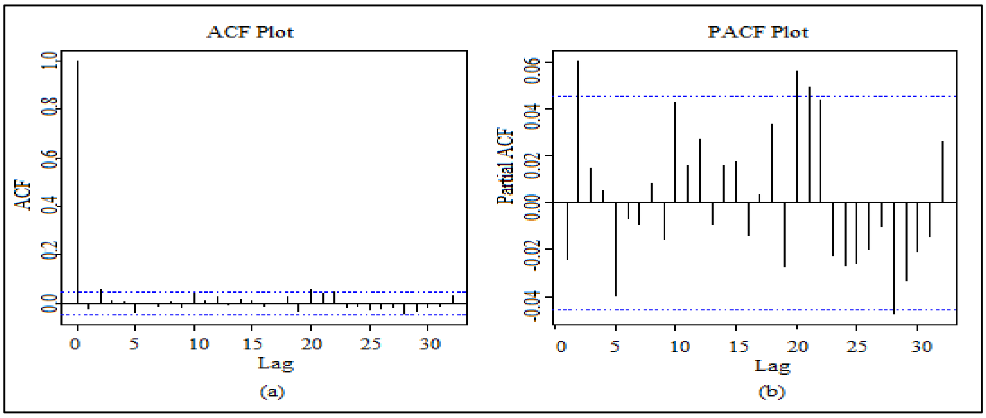


Figure 5: (a) The ACF plot and (b) PACF plot

The ACF plot, as seen in Figure 5, exhibits a minor increase at lag 2 with a lag value of 0.061 and no discernible increase after that. The diagram in the PACF plot depicts a sinusoidal pattern with one peak associated with hysteresis, two peaks at hysteresis 20 and 21, and no further discernible peaks. In each plot, but not in the first offset, it is possible to overlook a point that is out of limits if it is far from the first point. In spite of the fact that each graph contains 32 peaks, only roughly 30 percent of peaks are meaningful by coincidence. The shape of the second rise is consistent with that predicted for ARIMA because the PACF plot tends to degrade exponentially and the ACF has a significant 2 lag but not beyond that (0,1,2). Additionally, there is a likelihood that the ARIMA (0,1,2), ARIMA (0,1,3), and ARIMA (2,1,2) models will all fit the data well.

Table 4: The summary of ARIMA models

Model	(AIC)	(AIC_c)	(BIC)
ARIMA (0,1,2)	-9049.76	-9049.75	-9033.18
ARIMA (0,1,3)	-9048.41	-9048.39	-9026.30
ARIMA (2,1,2)	-9046.94	-9046.91	-9019.30

Table 4 shows that the ARIMA (0,1,2) model has the lowest values of the Akaike's Information Criterion (AIC), Modified Akaike's Information Criterion (*AICc*), and Bayesian Information Criterion when compared to the other two models (BIC). As a result, the top model to predict the daily stock price of FGV for the following 59 days will be the ARIMA (0,1,2) model.

```
Forecast method: Holt's method

Model Information:
Holt's method

Call:
 holt(y = fgv.train, h = 59)

  Smoothing parameters:
    alpha = 0.975
    beta  = 1e-04

  Initial states:
    l = 2.472
    b = -7e-04
```

Figure 6: The R software output of Holt linear trend method

The result of the R programmer is shown in Figure 6, where the proper smoothing parameter for the level represented by the alpha value is 0.975, the beta value reflects a trend of 1.000, the level's beginning state is 2.472, and the initial trend value is -0. 1282. In order to forecast the price of the FGV stock, smoothing factors for level and trend are utilised.

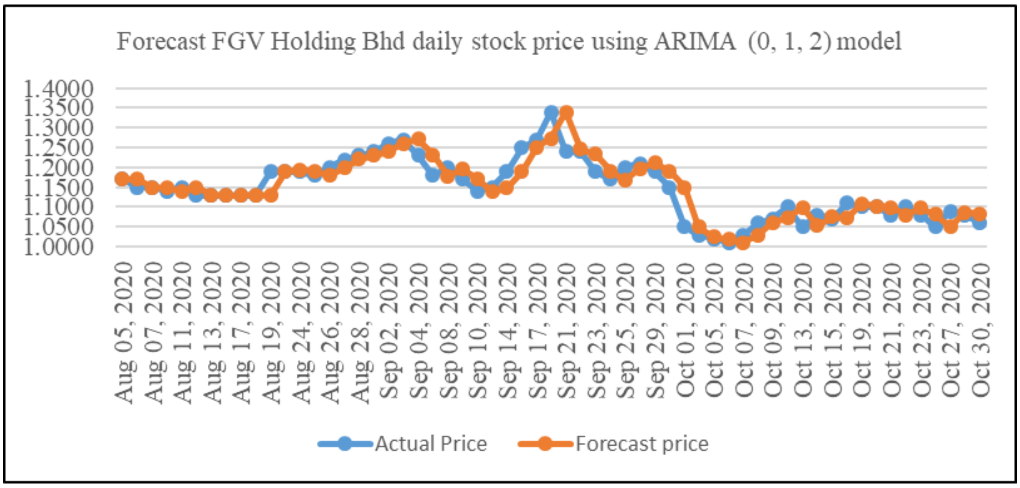


Figure 7: Comparison between the actual price and forecast price by using ARIMA (0, 1, 2)

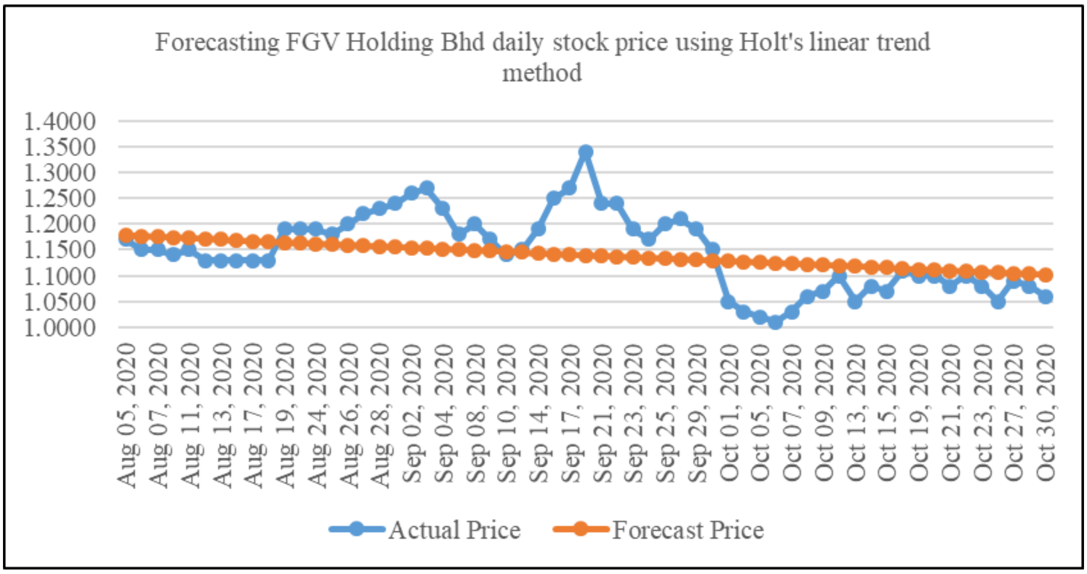


Figure 8: Comparison between the actual price and forecast price by using Holt's linear trend

The graph in Figure 7 illustrates how well the FGV stock price forecast made using ARIMA (0, 1, 2) matches the actual stock value. However, the stock price predicted using Holt's linear trend technique in Figure 8 does not exactly match the downtrend's actual stock price.

Table 5: The summary forecasting models applied

Forecasting model	MAE	MAPE	MSE	RMSE
ARIMA (0,1,2)	0.0235	0.0204	0.0010	0.0320
Holt's linear trend	0.0523	0.0451	0.0042	0.0647

Based on the summary from Table 5, shows ARIMA (0,1,2) model had lower forecast accuracy measures compared to Holt's linear trend method. As shown, in Table 5, ARIMA (0,1,2) had a lower mean absolute error which 0.0235 compared to Holt's linear trend method which was 0.0523. Therefore, it can be concluded that ARIMA (0,1,2) is the selected model to be an appropriate model to forecast the daily stock price for FGV due to the best forecasting method being the one that yields the lowest MAE, MAPE, MSE, and RMSE.

4. Conclusion

This study applied two methods to predict the Felda Global Venture Holdings Bhd (FGV) stock price from 2nd January 2013 until 30th October 2020 by using the forecasting methods which are ARIMA and Holt's linear trend method. This study explained that these two methods were relevant to predict the daily stock price and successfully achieve the objectives. The result of this study revealed that the ARIMA (0, 1, 2) was given the smallest forecasting error in predicting FGV stock price with the lowest error in MAE, MAPE, MSE and RMSE compared to Holt's linear trend method. Thus, the model of ARIMA (0, 1, 2) would be the best forecasting method for FGV stock price prediction and more upcoming plans can be created for increasing the stock price in the future since this model offers a good prediction accuracy for forecasting purpose. Application of the advanced forecasting methods is recommended to further this study due to it can enhance the predicting analysis and demonstrates better results with the lowest error. Moreover, this study also recommends adding more data due to the highest number of samples, the more accurate results will be performed for further analysis.

Acknowledgment

This research was made possible by funding from research grant TIER1 (Vot: H846) provided by Universiti Tun Hussein Onn Malaysia and the Ministry of Higher Education, Malaysia. The authors would like to thank the Faculty of Applied Sciences and Technology, Universiti Tun Hussein Onn Malaysia for the facilities provided that make the research possible.

References

[1] F. Ruhani, and S. Lecturer, "Review of the Literatures on Stock Price behaviour of Malaysia". International Journal of Islamic Business & Management, vol. 2, no. 2, pp. 32–38, 2018.

[2] FGV Holding Berhad (800165-P) "Company Overview FGV is one of the world's largest producers of crude palm oil (CPO), accounting for about 15% of Malaysia's total annual production of CPO". [Online] Available: https://www.fgvholdings.com/about-fgv/company-overview/, 2007 [Accessed February 16, 2020].

[3] Z. Jin, Y. Yang. and Y. Liu, "Stock closing price prediction based on sentiment analysis and LSTM". Neural Computing and Applications, 2, doi: https://doi.org/10.1007/s00521-019-04504-2, 2019.

[4] R. Hyndman, and G. Athanasopoulos, Forecasting Principle and Practice, Print Edition, August 2014.

[5] F. Sidqi, and I. D. Sumitra, "Forecasting Product Selling Using Single Exponential Smoothing and Double Exponential Smoothing Methods. IOP Conference Series: Materials Science and Engineering", 662(3), doi: https://doi.org/10.1088/1757-899X/662/3/032031, 2019.

[6] L. Ruekkasaem and M. Sasananan, "Forecasting Agricultural Products Prices Using Time Series Methods," 2018

[7] M. S. Arikan, M. B. ÇevrİmlĪ, B. Mat, and M. A. "Tekİndal Price Forecast for Farmed and Captured Trout Using Box-Jenkins", 2018.

[8] S. Çapar, "Importance of Initial Value in Exponential Smoothing Methods. Dokuz Eylül Üniversitesi Sosyal Bilimler Enstitüsü Dergisi,", vol. 17, no. 3, pp. 291, doi: 54 https://doi.org/10.16953/deusbed.12175, 2016.

[9] S. Prodanov, "Financial forecasting in the budgeting process", 2019

[10] F. K. Arya, and L. Zhang, "Time series analysis of water quality parameters at Stillaguamish River using order series method. Stochastic Environmental Research and Risk Assessment, vol. 29, no. 1, pp. 227–239, doi: https://doi.org/10.1007/s00477-014-0907-2, 2015.