

THE ACCURACY OF JUMP-DIFFUSION MODEL AND REGRESSION ANALYSIS FOR PREDICTION OF MALAYSIAN CENTRIFUGED LATEX PRICES

ZAWIN NAJAH HAMDAN, NUR HAIZUM ABD RAHMAN¹, AND SITI NUR IQMAL IBRAHIM

ABSTRACT. In this study, centrifuged latex prices in the Malaysian market are simulated using geometric Brownian motion with jump-diffusion process (GBMJ) and time series regression analysis. The sample daily data is collected from Malaysian Rubber Board over the period of 12 months. The centrifuged latex prices simulated using GBJM are compared with actual prices, prices simulated using geometric Brownian motion (GBM), and time series regression. Results show that over a short time interval, between GBM and GBJM, GBJM is more accurate than the GBM. However, between GBJM and time series regression, the latter is more accurate over all time range tested in this study.

1. INTRODUCTION

Simulation process involves at least approximately a set of rules that can be derived from historical prices, which implies realistic predictions if the underlying model used is realistic. Prediction of stock markets has been the interest in computational finance since investors are generally more interested in the changes in the stock price, than in its price itself.

Time series is one of the approach in statistics as a tool for forecasting. The historical observations are analyzed to develop a model that describes the relationship between variable and time. Then, the model is used to extrapolate the

¹*corresponding author*

2020 *Mathematics Subject Classification.* 60G22, 91G15.

Key words and phrases. Forecasting, Brownian motion, centrifuged latex prices, jump-diffusion, financial performance, short term investment.

series into future. Time series methods can be classified into two; classical and modern [1]. However, there is no clear-cut evidence shows that the modern methods can consistently outperform the classical methods in forecasting for all situations. Researcher tend to use classical methods due to its simplicity and acceptable forecast accuracy [2]. On the other hand, the modern methods are approach which promise the good results in many forecasting areas. This is because modern method appropriate when the situation of the problem is unclear [3].

In financial market, uncertain movements of asset prices reflect its dynamics over time. According to [4], future stock prices depend on current stock prices following the efficient market hypothesis (EMH), which explains the randomness behavior of stock prices. Thus, modeling asset prices is crucial to model the new information about the stock. This is because the historical prices of an asset does not say a lot about the behavior of its future price. On that account, this implies that the future prices of an asset can be predicted given its current price. This randomness is described by a financial process with the assumption of normally distributed and independent stock returns in modeling stock prices, namely the geometric Brownian motion (GBM). The ability of GBM to simulate stock price paths for a short time interval attracts investors who wants to gain profit in a short-term investment. The potential of GBM as an effective forecasting method has been studied to forecast the future closing prices for small sized companies listed in Bursa Malaysia.

The GBM process which assumes the stock returns are independent and normally distributed. However, the distribution of the logarithmic stock price is negatively skewed due to large fluctuations, such as crashes, and evidence of price jumps or shocks, exhibit by the stock price. Several studies have presented strong empirical evidence in favour of jumps in financial assets, which also discredits the classical continuous time paradigm with continuous sample price paths inherent in the use of governing the asset price process solely by a Brownian motion process. For instance, [5] investigates sample paths of exchange rates and stock market index and proves existence of discontinuities. A seminal work of [6] applies a GBM model which incorporates Poisson jumps to approximate the path of stock prices subject to occasional shocks, hence capturing the negative skewness and excess kurtosis of the logarithmic stock price density.

Other than the stock prices, commodity prices may also exhibit mean-reversion and jumps [7]. The jump process is assumed to be independent of the commodity

price process. We study the accuracy of the GBM with jumps (GBMJ) and time series regression analysis in simulating Malaysian centrifuged latex price paths. The remainder of the paper is organized as follows. In Section 2, we review the GBM, GBMJ and time series regression used in this study. Section 3 documents the numerical results with illustrations, and Section 4 concludes the paper.

2. MATERIALS AND METHODS

In this section, we describe the models that are used in the study, which are the geometric Brownian motion (GBM), GBM with jump-diffusion process (GBMJ), and time series regression analysis. Following [8], the behavior of the rubber prices P_t is modeled as a GBM process as such:

$$(2.1) \quad dP_t = \mu P_t dt + \sigma P_t dW_t,$$

which can be represented in terms of proportional returns P_t as follows:

$$\frac{dP_t}{P_t} = \mu dt + \sigma dW_t,$$

where μ is the rate of return, σ is the volatility, and W_t is a standard Brownian motion that is normally distributed with mean zero and standard deviation of a square root of the time step. Using some basic stochastic calculus, (2.1) can be expressed as follows:

$$(2.2) \quad d(\ln P_t) = (\mu - \frac{1}{2}\sigma^2)dt + \sigma dW_t,$$

which implies $\ln P_t$ follows arithmetic Brownian motion dynamics with mean $\mu - (1/2)\sigma^2$ and variance $\sigma^2 dt$. Over a time interval $[t_i, t_{i-1}]$, integration on both sides of (2.2) yields the following expression:

$$P_{t_i} = P_{t_{i-1}} e^{(\mu - \frac{1}{2}\sigma^2)(t_i - t_{i-1}) + \sigma(W_{t_i} - W_{t_{i-1}})},$$

which is the GBM model of the future commodity price paths that follows a log-normal distribution. The Euler discretization of the log commodity price process is:

$$P_{t_i} = P_{t_{i-1}} e^{(\mu - \frac{1}{2}\sigma^2)\Delta t + \sigma\sqrt{\Delta t}Z_i},$$

where $Z_i \sim N(0, 1)$. Given historical prices, parameters μ and σ are estimated using historical logarithmic returns at fixed time intervals as follows, respectively

(see [8]):

$$\mu = \bar{R} = \frac{\sum_{i=1}^M R_{t_i}}{M},$$

where $R_{t_i} = \ln(P_{t_i}/P_{t_{i-1}})$, and

$$\sigma = \sqrt{\frac{\sum_{i=1}^M R_{t_i} - \bar{R}}{(M-1)}}.$$

This completes the GBM model.

In order to model the movement of commodity prices that exhibit jumps, we follow [6] model by adding a Poisson jump process to the GBM in (2.1) (GBMJ hereafter) as such:

$$(2.3) \quad dP_t = \mu P_t dt + \sigma P_t dW_t + P_t dN_t,$$

where W_t is the standard Brownian motion, σ is the volatility, and N_t is the univariate jump process that is defined as $N_T = \sum_{n=1}^{J_T} (Y_n - 1)$ where the $Y_n \sim \exp(N(\mu_Y, \sigma_Y^2))T$ which are i.i.d. log-normal variables, and the Poisson probability is given by:

$$\mathbf{P}(N(t) = k) = e^{-\lambda T} \frac{(\lambda T)^k}{k!},$$

for $n = 0, 1, 2, \dots$. In terms of the log price, (2.3) can be expressed as follows:

$$d(\ln P_t) = (\mu - \frac{1}{2}\sigma^2)dt + \sigma dW_t + \ln Y_{J_t} dJ_t.$$

Over a time interval $[t_i, t_{i-1}]$, integrating both sides of (2) yields the following expression:

$$P_{t_i} = P_{t_{i-1}} e^{(\mu - \frac{1}{2}\sigma^2)(t_i - t_{i-1}) + \sigma(W_{t_i} - W_{t_{i-1}})} \prod_{n=1}^{J_T} Y_n,$$

which is the GBMJ model of the future commodity price paths. The Euler discretization of the log commodity price process is:

$$P_{t_i} = P_{t_{i-1}} e^{(\mu - \frac{1}{2}\sigma^2)\Delta t + \sigma\sqrt{\Delta t_i}Z_i} \prod_{n=1}^{J_T} Y_n,$$

where $Z_i \sim N(0, 1)$ and the jumps between the time interval is $j_{t_i} = J_{t_i} - J_{t_{i-1}}$.

Given the logarithmic returns conditioning on jumps, the model parameters $\mu^*, \mu_j, \sigma^2 > 0, \sigma_j^2 > 0$ and $\lambda > 0$ are calibrated using maximum likelihood estimation (MLE) method as $\operatorname{argmax} L^* = \log L$.

The log-likelihood function for the returns P_t is given as follows:

$$L^*(\theta) = \sum_{i=1}^n \log f(x_i; \mu^*, \mu_j, \sigma^2, \sigma_j^2, \lambda)$$

for $t = t_1, \dots, t_n$ where $\Delta t = t_i - t_{i-1}$ and $f(\cdot)$ is the probability density function as such:

$$(2.4) \quad f(x_i; \mu^*, \mu_j, \sigma^2, \sigma_j^2, \lambda) = \sum_{j=0}^{\infty} P(n_t = j) \times f_N[x_i; (\mu - \frac{1}{2}\sigma^2)\Delta t + j\mu_j, \sigma^2\Delta t + j\sigma_j^2]$$

for $x_1, \dots, x_n$. (2.4) is an infinite mixture of Gaussian random variable, each weighted by a Poisson probability $P(n_t = j) = f_p(j; \lambda\Delta t)$. Following that, (2.4) can be simplified to a mixture of two Gaussian random variable weighted by the probability of zero or one jump in Δt as follows:

$$f(x_i; \mu^*, \mu_j, \sigma^2, \sigma_j^2, \lambda) = (1 - \lambda\Delta t)(f_N[x_i; (\mu - \frac{1}{2}\sigma^2)\Delta t + \sigma^2\Delta t]) + \lambda\Delta t(f_N[x_i; (\mu - \frac{1}{2}\sigma^2)\Delta t + \mu_j, \sigma^2\Delta t + \sigma_j^2]).$$

This completes the GBMJ model.

Meanwhile, regression analysis is one of statistical tools in estimating the relationships between variables. By measuring the random variable, y to another variable of x . One of the main regression analyses is simple linear regression where the variables linear in the parameters. A simple linear regression model can be written as $Y_i = \alpha + \beta X_i + \epsilon_i$, where Y is a random variable, X is a fixed variable and ϵ is a random error term whose value is based on an underlying probability distribution usually normal distribution.

The statistical procedure in finding the best fitting line for a set of bivariate data that minimized the distance or deviations can be formulated as:

$$y = \hat{\alpha} + \hat{\beta}x,$$

where a and b are the estimates of the intercept and slope parameters of α and β , respectively. Ordinary least square (OLS) is the method to estimate the parameters model. OLS method assumes that the residual of the model satisfies white noise and normally distributed condition (Mendenhall et al., 2012). The least square

method objective is to find a line of best fit, so the error is minimized:

$$\sum_{i=1}^n \hat{\epsilon}_i^2 = \sum_{i=1}^n (Y_i - \hat{y}_i)^2.$$

To minimize the sum of squared residuals, taking the partial derivative with respect to α and β , setting each derivative equal to 0 and solving these derivatives simultaneously. Thus, the estimate parameter of α and β can be written as:

$$\hat{\beta} = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2}$$

and

$$\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X}.$$

In the case of time series analysis, the time series regression models been used. These models relate the dependent variable y_t to functions of time, t . The model is useful when the parameters describing the series remain constant over time. For example, linear trend series where the slope remain constant over time. The time series, y_t by using a trend model can be written as:

$$y_t = TR_t + \epsilon_t,$$

where y_t is the value of the time series in period t , TR_t is the trend in time period t , ϵ_t is the error term in time period t .

The TR_t can be described into several forms, which are no trend where short growth or decline, $TR_t = \beta_0$, linear trend is a model with long growth or long decline, $TR_t = \beta_0 + \beta_1 t$, and quadratic trend is a quadratic with long run change over time, $TR_t = \beta_0 + \beta_1 t + \beta_2 t^2$.

This completes the regression model.

3. RESULTS AND DISCUSSION

This section illustrates the commodity price path simulation using GBMJ and time series regression analysis. We use 244 daily closing prices of Malaysian centrifuged latex at noon from 2 December 2013 until 28 November 2014, which are obtained from the official website of Malaysian Rubber Board [9]. In order to evaluate the accuracy of the prediction prices obtained from the models, we compute

the mean absolute percentage error (MAPE) that is defined as such:

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{P_t - F_t}{P_t} \right| \times 100\%.$$

Then, by referring to Lewis' judgement scale [10] as shown in Table 1 to determine how accurate the predictions are based on the MAPE values.

TABLE 1. Lewis Judgement Scale

MAPE Values	Rate of Accuracy
$\leq 10\%$	High
11% to 20%	Good
21% to 50%	Reasonable
$\geq 51\%$	Inaccurate

We simulated the future prices of the centrifuged latex for a range of different time period, which is over a period of one month to a period of 12 months, using the GBMJ and time series regression analysis.

TABLE 2. MAPE Values for Forecast Prices using GBMJ and Regression

Simulation Period	GBMJ	Regression
1 month	3.74	8.28
2 months	8.17	5.26
3 months	7.97	4.37
4 months	9.20	4.76
5 months	9.93	4.81
6 months	11.12	4.58
7 months	13.89	4.90
8 months	17.62	5.19
9 months	23.57	5.03
10 months	29.51	4.77
11 months	34.53	4.69
12 months	36.94	4.66

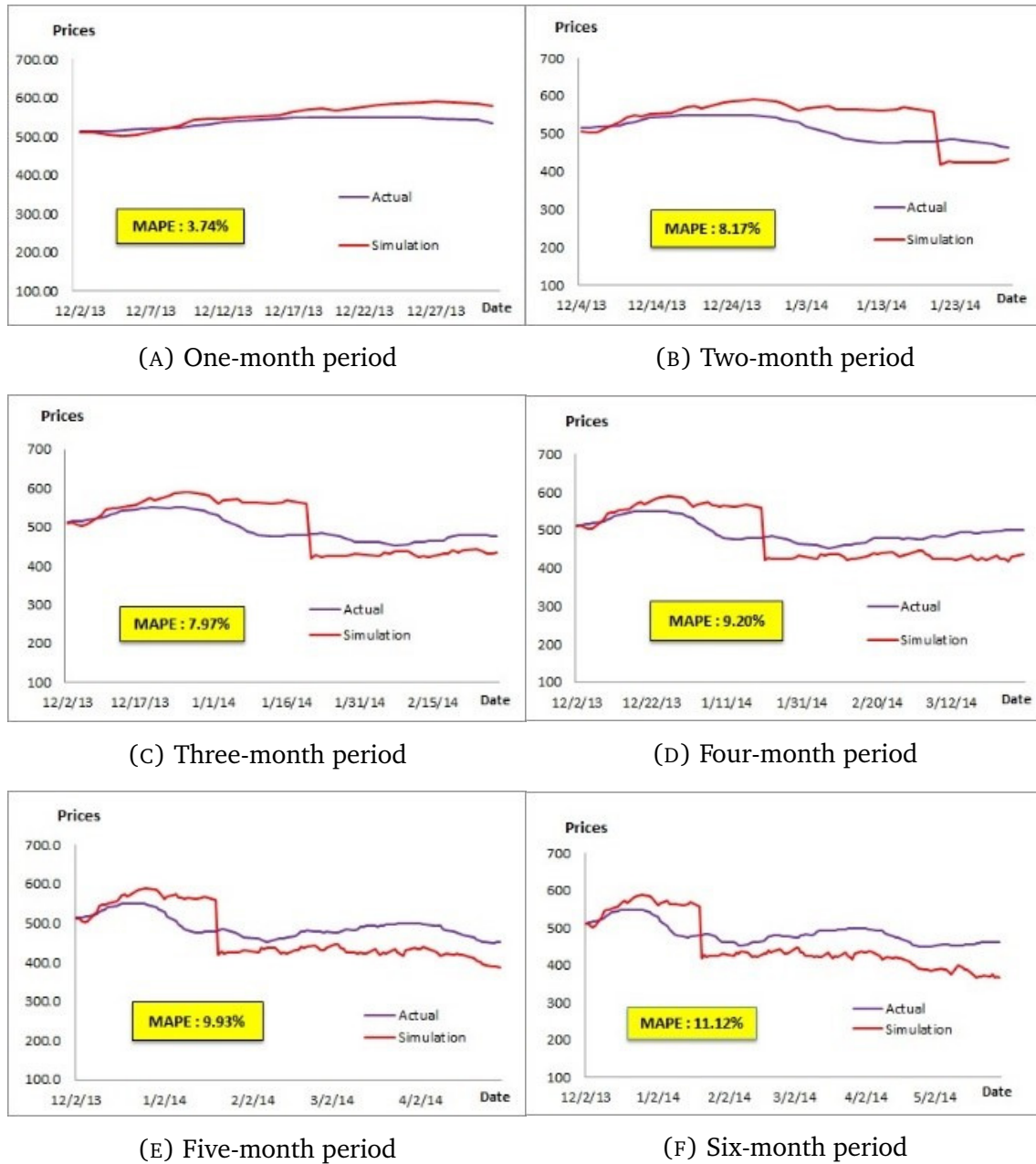


FIGURE 1. Actual vs GBMJ Forecast Prices over 1-6-month period

Figure 1 and 2 plot the actual and prediction prices for twelve different time intervals, and Table 2 tabulates the MAPE values. It can be seen from Table 2 that GBMJ produces highly accurate prediction prices over a one-month period

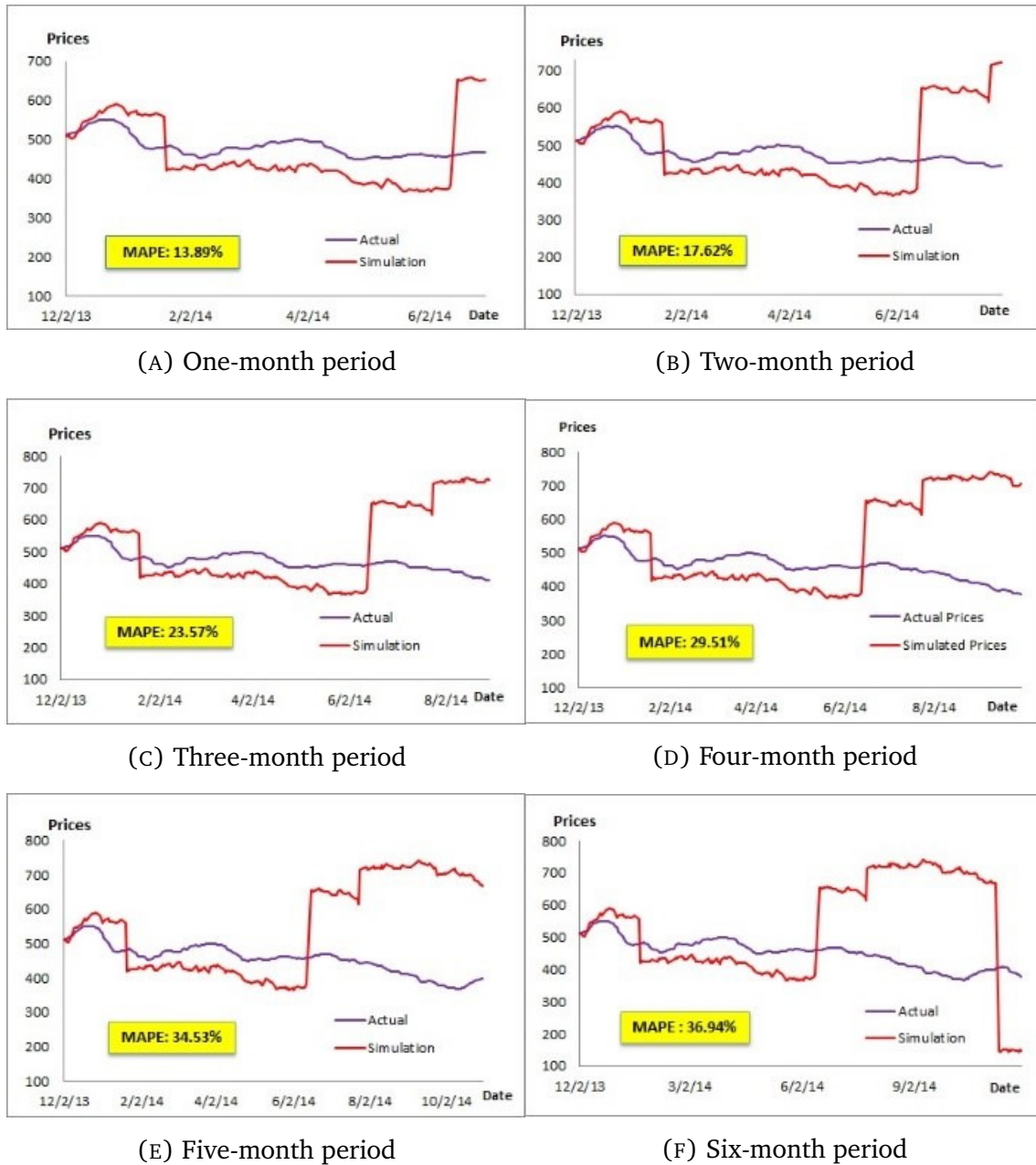


FIGURE 2. Actual vs GBMJ Forecast Prices over 7-12-month period

up until a period of five months, during which the accuracy is the highest for a one-month period. The difference in the MAPE values from a period of one month to five is quite significant. This can also be observed from Figures 1 and 2 that

TABLE 3. Regression Summary

Predictor	Coefficient	Standard Error	Coefficient	T	P
Constant	730.265	3.988		185.43	0.000
Time	-0.498 2	0.0145		-34.31	0.000

TABLE 4. Analysis of Variance (ANOVA) Result

Source	DF	Sum of Square	Mean Square	F	P
Regression	1	2134030	2134030	1171.30	0.000
Residual Error	467	846505	1813		
Total	468	2980535			

the prediction prices are most aligned to the actual prices for a month period (see Figure 1(a)) compare to the other eleven time periods (see Figures 1(b)-2(f)). As the time interval increases from one month to twelve months, the accuracy rate for GBMJ decreases. However, the regression analysis produces highly accurate forecast for all time range, even for a longer time period. This shows that time series regression analysis outperforms GBMJ.

For comparison, Figure 3 plots the prediction prices for a one-month period via GBM and GBMJ. On the other hand, Figure 4 shows the output for latex prices data by using regression method. Given the constant value is 730.265 while the intercept value is -0.498. This result indicates that the centrifuged latex prices decreases as time increases.

In order to examine either the model is significant or not significant the hypothesis testing is conducted by using $H_0 : \beta = 0$. Based on the p-value ($p=0.000$) of t-statistic in Table 3, H_0 is rejected which indicates a significant linear relationship between explanatory variable and response variable ($\beta_0 \neq 0$) exists. Moreover, the analysis of variance (ANOVA) result in Table 4 shows a significant linear relationship since the p-value is 0.000. Additionally, the strength of the relationship between time and centrifuged latex prices can be measured using the coefficient of determination, R^2 , which is 71.6%; thus implies the effectiveness of the regression model.

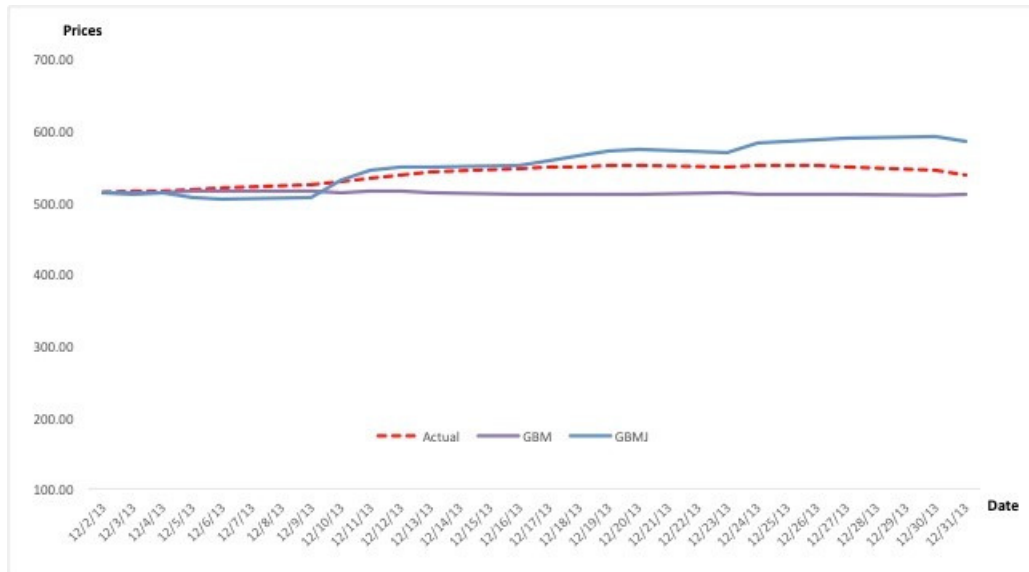


FIGURE 3. Actual Price, GBM Forecast Prices and GBMJ Forecast Prices over a One-Month Period

4. CONCLUSION

This study applies a mathematical approach to predict future prices of centrifuged latex in Malaysian market. The advantages of geometric Brownian motion (GBM) involves a more straightforward computation and the short duration of data used is sufficient to predict future prices of stocks or commodities short-term. This study has shown that by incorporating jumps to the GBM model that describes the market as close to the actual market, the accuracy of the prediction prices increases for short time interval. Hence investors can use GBMJ to predict future prices of centrifuged latex for short-term investment because the accuracy of the prediction deteriorates as the number of time periods increases. Nevertheless, for a long-term investment, the time series regression analysis shows highly accurate predictions.

Furthermore, the component in time series give important information before proceeding into the analysis steps. Based on the centrifuged latex price, it shows decline pattern (trend). Besides time series regression, there are several other methods that can be used for analysis, such as the exponential smoothing with trend and autoregressive moving average model (ARIMA), which is widely used in time series application due to its capability in modelling different type of series

with success. In addition, these models also important in setting the benchmark method in comparing to the new develop method for modelling and forecasting time series data with great accuracies. Future work can include incorporating other factors such as seasonality to GBM or GBMJ to capture the movement of the price paths.

ACKNOWLEDGMENT

The author acknowledges the support given by Universiti Putra Malaysia under project number GP-IPM/2017/9587700.

REFERENCES

- [1] M. ELENA, M.H. LEE, SUHARTONO, HOSSEIN, N. HAIZUM, N.A. BAZILAH: *Fuzzy time series and sarima model for forecasting tourist arrivals to Bali*, J, Teknol., **57**(1) (2012), 69-81.
- [2] R.J. HYNDMAN, G. ATHANASOPOULOS: *Forecasting: Principles and Practice: Notes*, 1st ed., OTexts, Australia, 2014.
- [3] SUHARTONO, D.E. ASHARI, D.D. PRASTYO, H. KUSWANTO, M.H. LEE: *Deep neural network for forecasting inflow and outflow in Indonesia*, Sains Malays., **48**(8) (2019), 1787–1798.
- [4] E.F. FAMA: *Efficient Capital Markets: A Review of Theory and Empirical Work*, J. Finance., **25**(2) (1970), 383-417.
- [5] P. JORION: *On jump processes in the foreign exchange and stock markets*, Rev Financ Stud., **1**(4) (1988), 427-445.
- [6] R. MERTON: *Option pricing when underlying stock returns are discontinuous*, J. Financ Econ., **3**(1-2) (1976), 125-144.
- [7] W. CAO, S.B. GUERNSEY, S.C. LINN: *Evidence of infinite and finite jump processes in commodity futures prices: Crude oil and natural gas*, Physica A., **502** (2018), 629-641.
- [8] S.Z. ABIDIN, M.M. JAFFAR: *Forecasting Share Prices of Small Size Companies in Bursa Malaysia Using Geometric Brownian Motion*, Appl. Math. Inform. Sci., **8**(2) (2014), 107-112.
- [9] MALAYSIAN RUBBER BOARD: *Monthly reference prices for physical rubber (free on board) SMR grades and centrifuged latex (60% DRC) in sen/kg[Data file]*. Retrieved from: <http://www3.lgm.gov.my/mre/MonthlyPrices.aspx>, 2020.
- [10] K.D. LAWRENCE, R.K. KLIMBERG, S.M. LAWRENCE : *Fundamentals of Forecasting using Excel*, Industrial Press, New York, 2009.

DEPARTMENT OF MATHEMATICS, FACULTY OF SCIENCE,
UNIVERSITI PUTRA MALAYSIA,
43400 UPM SERDANG, SELANGOR, MALAYSIA.
Email address: awin.hamdan@gmail.com

DEPARTMENT OF MATHEMATICS, FACULTY OF SCIENCE,
UNIVERSITI PUTRA MALAYSIA,
43400 UPM SERDANG, SELANGOR, MALAYSIA.
Email address: nurhaizum_ar@upm.edu.my

INSTITUTE FOR MATHEMATICAL RESEARCH & DEPARTMENT OF MATHEMATICS,
FACULTY OF SCIENCE, UNIVERSITI PUTRA MALAYSIA,
43400 UPM SERDANG, SELANGOR, MALAYSIA.
Email address: iqmal@upm.edu.my