JOURNAL OF INFORMATION SCIENCE AND ENGINEERING 40, 1045-1056 (2024)
DOI: 10.6688/JISE.202409_40(5).0008

Comparing Model Building Performance of ARIMA Model and Logarithmic Return Model

SIAW LI LEE+, CHIN YING LIEW, CHEE KHIUM CHEN AND LI LI VOON
Department of Mathematical Sciences
College of Computing, Informatics and Mathematics
Universiti Teknologi MARA, Sarawak Branch
Kota Samarahan, 94300 Malaysia
E-mail: siawli@uitm.edu.my+; cyliew@uitm.edu.my; ckchen@uitm.edu.my; llvoon@uitm.edu.my

Autocorrelated statistical process control that is widely employed in process control environments typically uses the autoregressive integrated moving average (ARIMA) model in fitting autocorrelated time series data. Nevertheless, the iterative modeling procedures of ARIMA are laborious, time-consuming, expensive, and complex. Meanwhile, autocorrelated data is governed by the geometric Brownian motion (GBM) law if its logarithmic returns are independent and identically normally distributed (i.i.n.d.). By utilizing these attributes, this paper aims to propose the Logarithmic Return (LR) model as an alternative methodology in modeling time series data. Twelve real-world datasets are used to demonstrate the applicability of the proposed model. All computations are implemented via R-programming language. In addition to being parsimonious and easy to compute, the LR model is reported with a shorter Central Processing Unit (CPU) running time. Specifically, it typically takes an average of less than 0.20 seconds to obtain the LR model using twelve datasets, while its counterpart requires over 5 seconds. LR model has a comparable good mean average percentage error (MAPE) to the ARIMA model, thus LR model is as accurate as the ARIMA model. This study shows that the LR model with two parameters and requires a two-step implementation procedure is a promising alternative model of ARIMA for positive datasets in time series modeling.

Keywords: ARIMA, autocorrelated statistical process control, geometric Brownian motion, logarithmic return, parsimonious

1. INTRODUCTION

Statistical methods have been broadly applied in process control environments to evaluate, improve, and control quality characteristics, and to reduce the variability process [1-3]. Among the methods, control charts are often utilized in manufacturing processes to monitor continuous processes as part of its quality control to lessen the risk of loss [4]. The use of a control chart presumes each random observation in the samples as mutually independent and identically normally distributed (i.i.n.d.), but autocorrelated observations are common in practice [5] due to frequent measurements taken at short time intervals [6]. Autocorrelation usually appears in manufacturing processes [7]. To prevent false signals from occurring in control charts that can lead to undesired problems, autocorrelated statistical process control (SPC) techniques work to eliminate autocorrelation. One of the typical techniques employed in autocorrelated SPC is to fit a time series model to autocorrelated data and use the i.i.n.d. residuals to generate control charts [5, 8-12].

Studies show that the autoregressive integrated moving average (ARIMA) model is

Received February 3, 2023; accepted October 18, 2023.
Communicated by Beng Yong Lee.

widely accepted to fit autocorrelated time series data, thus chosen as the benchmark for the present study. It is also called the Box-Jenkins's methodology [13] and seen in industrial production [14] and manufacturing processes [15]. ARIMA is regarded as a stable statistical technique that gives efficient and accurate solutions. However, researchers commented that the tough process to estimate the parameters, and validate the identified model are time wasting, expensive, and complex [16-18]. Specifically, the challenges include the ambiguity plots of autocorrelation function (ACF) and partial autocorrelation function (PACF) and the overfitting problem [13, 19, 20]. Moreover, researchers stipulated that statistical modeling should aim for a simple best fit model that can still explain the data [21]. This has motivated the current study to work out a parsimonious model that potentially overcomes the difficulties encountered in ARIMA modeling.

This study aims to propose a time series model to fit autocorrelated data and compare its model building performance in terms of mean average percentage error (MAPE) and running time with the standard practice, ARIMA modeling. This paper is an extension of work published in [22], which utilizes the attributes of geometric Brownian motion (GBM) law to propose the Logarithmic Return (LR) model as an alternative methodology in modeling positive time series autocorrelated data. An autocorrelated data is governed by the GBM law if the logarithmic returns of the continuous process are i.i.n.d. The logarithmic returns are derived by taking appropriate logarithm transformation of the autocorrelated data, which gives the LR model and fulfil the independence and normality assumption of residuals.

The preliminary evidence on the benefits of LR model is further studied on several real-world datasets. In this paper, LR and ARIMA models were applied on eleven secondary data and one industrial data where the residuals were evaluated for its accuracy and speed. These secondary datasets comprise the environmental, consumption, population, economics, stock fund, services, sales, and chemical processes, while the industrial data is collected by the researchers during a manufacturing process. The results showed that LR model is parsimonious and can be computed easily, reduces CPU running time, and has comparable good MAPE to the ARIMA model. The proposed methodology with LR model is presented next. This is followed by the discussions of results obtained by the ARIMA and LR models for the twelve datasets, and the comparison of both models. Lastly, the paper ends with a conclusion.

2. PROPOSED LOGARITHMIC RETURN (LR) MODEL

In this section, a new procedure of time series modeling corresponding to the LR model is proposed. See [22] for the details of LR model derivation. Let X_t be a positive and continuous time series. The theoretical procedure consists of the following processes:

i) test the independency (presence of autocorrelation) of X_t,
ii) transform X_t into $R_t = \ln(X_t) - \ln(X_{t-1})$,
iii) test the normality of the R_t,
iv) test the independency (absence of autocorrelation) of R_t,
v) compute the parameters in the first order autoregressive model of R_t and the LR model is,

$$\hat{X}_t = e^{\hat{C}} \cdot X_{t-1} \cdot \left(\frac{X_{t-1}}{X_{t-2}} \right)^{\hat{\Theta}} \quad (1)$$

vi) use the model in Eq. (1) to describe the data and conduct further analysis if its prediction accuracy is as expected. To evaluate the prediction accuracy of the model, the MAPE is calculated,

$$\text{MAPE} = \frac{1}{m} \sum_{t=1}^{m} \left| \frac{X_t - \hat{X}_t}{X_t} \right| \times 100\%$$

Here, X_t is the actual value, $\hat{X}_t$ is the predicted value given by Eq. (1), and m is the number of all residuals. As a rule of thumb, the meaning of MAPE is given in Table 1 [23].

Table 1. MAPE and prediction accuracy.

MAPE	Prediction Accuracy
< 10%	Highly accurate forecast
11% to 20%	Accurate forecast
21% to 50%	Reasonable forecast
> 50%	Inaccurate forecast

3. RESULTS AND DISCUSSION

3.1 Exploratory Data

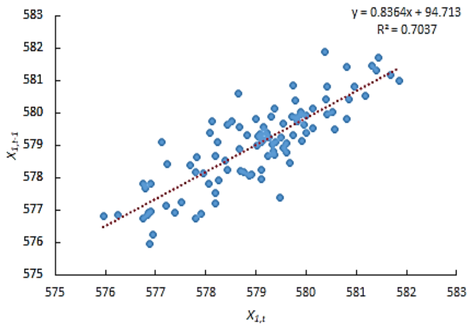
A great number of positive datasets available in the literatures and websites were investigated. Due to the limited space, eleven examples of public data source and one industrial data are used to illustrate the advantages of the proposed model. The data used in this study are shown in Table 2. Data 12 was collected by researchers at a cocoa powder factory for a period of two months. In a typical cocoa powder production, the color of cocoa powder is an essential quality characteristic that needs to be observed closely during the production process. The cocoa powder color attributes L^*, a^* and b^* were measured using the ColorFlex EZ meter. The value L^* indicates lightness, a^* denotes the red-green variation while b^* denotes the yellow-blue variation. If no color correction is done, it will increase the waste in fermentation, drying and roasting process. Moreover, it would impact the end product and vary in color from light brown to dark red or black. Therefore, a careful and rigorous SPC on color is necessary in food industry. All computation is conducted by using *R*-programming language. For each dataset, both ARIMA and LR models are constructed.

3.2 ARIMA Models

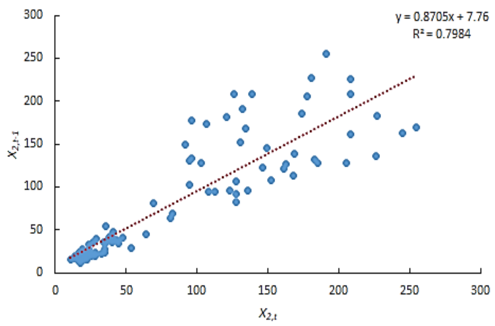
In this paper, twelve time series datasets are used. The presence of the autocorrelation in the data can be visualized through the lag-1 scatter plot in Fig. 1. From these figures, it was clearly shown that the process data was autocorrelated.

Table 2. Description of twelve datasets.

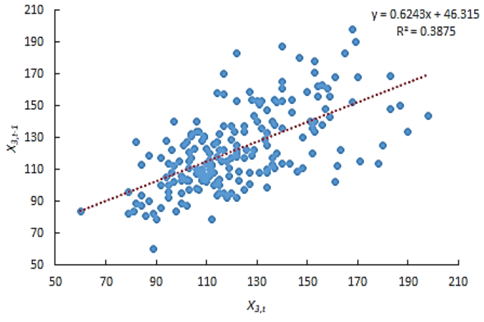
Dataset	Description
Data 1	The water level of Lake Huron from 1875 to 1972 [24].
Data 2	U.K. coal and other solid fuels consumption from 1960 to 1986 [25].
Data 3	The number of car drivers killed in Great Britain from January 1969 to December 1984 [25].
Data 4	The number of muskrat furs traded annually by the Hudson's Bay Company in Canada from 1848 to 1909 [25].
Data 5	Australian imports goods and services at average 1984/5 prices [26].
Data 6	Monthly U.S. beer sales from January 1975 to December 1990 [27].
Data 7	Monthly public transit boarding in Denver, Colorado, region from August 2000 to March 2006 [27].
Data 8	Daily values of one unit of the College Retirement Equity Fund (CREF) stock fund from August 26, 2004 to August 15, 2006 [27].
Data 9	Monthly unit sales of recreational vehicles from Winnebago, Inc. from November 1966 to February 1972 [27].
Data 10	The chemical process viscosity readings [28].
Data 11	U.S. beverage manufacturer product shipments from January 1992 to December 2006 [28].
Data 12	The color reading of cocoa powder production process at cocoa powder industry (located in Johor Bahru, Malaysia) from June 2011 until July 2011.



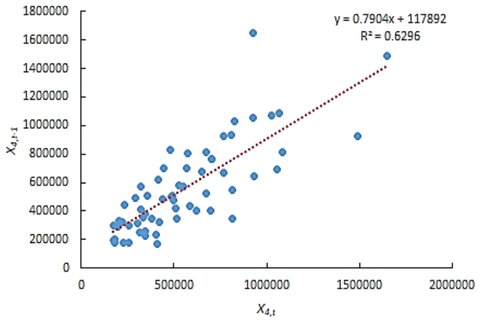
(a) Data 1



(b) Data 2



(c) Data 3



(d) Data 4

Fig. 1. Lag-1 scatter plot of twelve datasets.

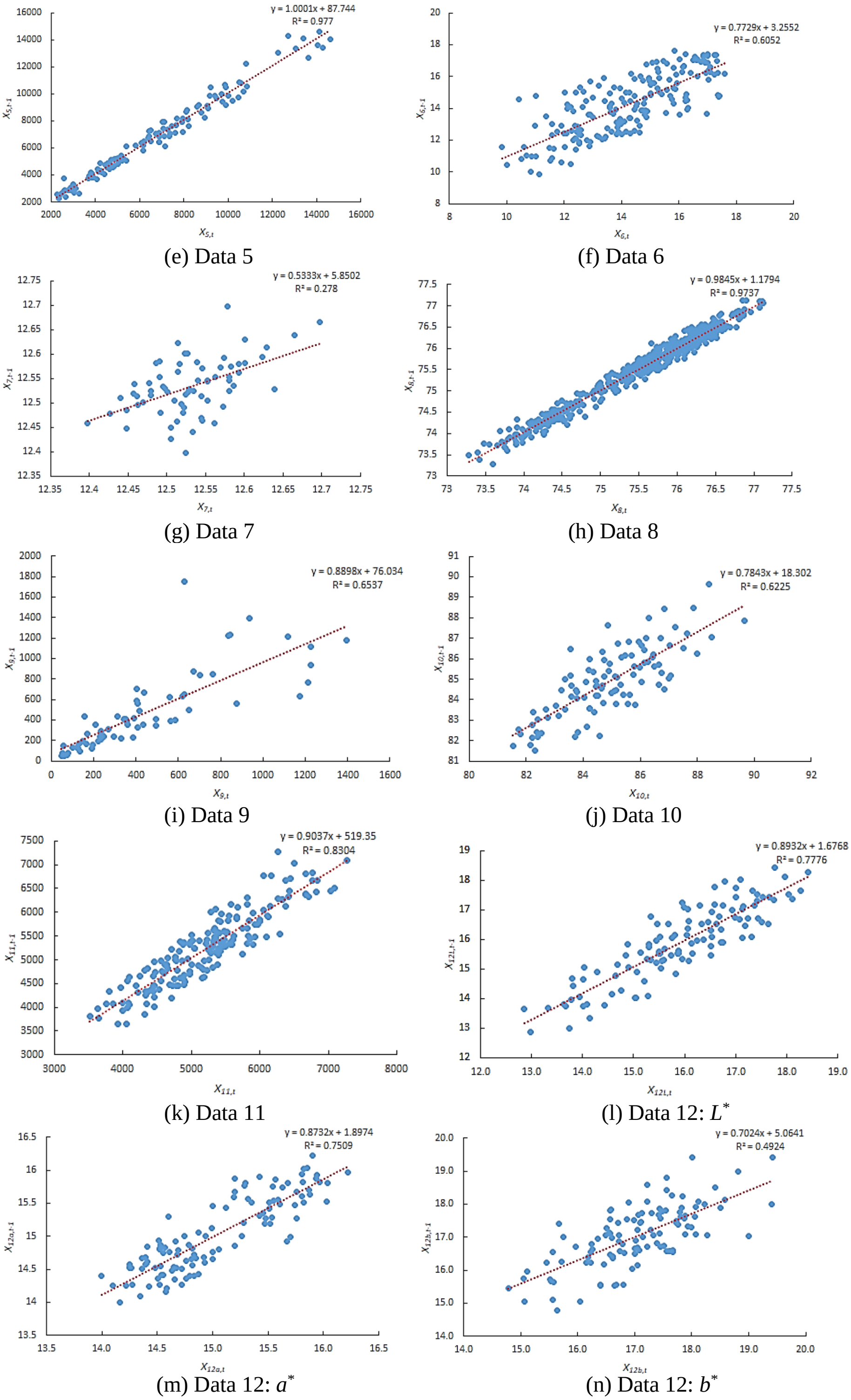


Fig. 1. (Cont'd) Lag-1 scatter plot of twelve datasets.

For each dataset, the best ARIMA model is determined by using *R*-programming language and the accuracy of each model based on MAPE is calculated as shown in Table 3. Referring to the table, most of the models are at the high accuracy level with MAPE values between 0.101134% and 5.894200%. ARIMA models of Datas 2 and 3 are at the accurate level with MAPE 19.099914% and 12.446644%, respectively, while Data 4 and Data 9 show reasonable accuracy with MAPE 25.161284% and 34.66659%, respectively.

Table 3. ARIMA model for twelve datasets.

Dataset	Best model	ARIMA Model	MAPE (%)
Data 1	ARIMA (0,1,0)	$\hat{X}_t = X_{t-1}$	0.101134
Data 2	ARIMA (0,1,4)	$\hat{X}_t = X_{t-1} - 0.2011593e_{t-1} - 0.6374598e_{t-2} + 0.2020225e_{t-3} + 0.5539320e_{t-4}$	19.099914
Data 3	ARIMA (2,1,2)	$\hat{X}_t = 2.4895842X_{t-1} - 2.2012217X_{t-2} + 0.7116375\,X_{t-3} - 1.8346615e_{t-1} + 0.8995951e_{t-2}$	12.446644
Data 4	ARIMA (1,1,2)	$\hat{X}_t = 1.4744611X_{t-1} - 0.4744611X_{t-1} - 0.2832056e_{t-1} - 0.5900290e_{t-2}$	25.161284
Data 5	ARIMA (1,1,1)	$\hat{X}_t = 1.8179832X_{t-1} - 0.8179832\,X_{t-2} + 0.9337341e_{t-1} + 86.5637107$	5.219641
Data 6	ARIMA (1,1,3)	$\hat{X}_t = 0.5900919X_{t-1} + 0.4099081X_{t-2} + 3515607e_{t-1} + 0.3723849e_{t-2} + 0.6398156e_{t-3}$	5.894200
Data 7	ARIMA (3,1,2)	$\hat{X}_t = 1.6493246X_{t-1} - 1.2742097X_{t-2} + 0.3274908X_{t-3} + 0.2973943X_{t-4} + 1.1439950e_{t-1} + 0.7467929e_{t-2}$	0.255938
Data 8	ARIMA (2,1,2)	$\hat{X}_t = 1.1772687X_{t-1} - 2.1417163\,X_{t-2} + 0.9644476\,X_{t-3} - 1.2059162e_{t-1} - 0.9582186e_{t-2}$	0.143011
Data 9	ARIMA (0,1,1)	$\hat{X}_t = X_{t-1} - 0.6825704e_{t-1} + 19.6262288$	34.66659
Data 10	ARIMA (0,1,0)	$\hat{X}_t = X_{t-1}$	1.022171
Data 11	ARIMA (5,1,1)	$\hat{X}_t = 1.4428176X_{t-1} - 0.284376X_{t-2} - 0.3628299X_{t-3} - 0.7424045X_{t-4} + 0.3777078X_{t-5} - 0.1603084X_{t-6} - 0.8831909e_{t-1} - 14.0800381$	3.518228
Data 12: L^*	ARIMA (2,1,2)	$\hat{X}_t = 2.6400951X_{t-1} - 0.3525548X_{t-2} - 0.7684622\,X_{t-3} - 1.8995388e_{t-1} + 0.9672713e_{t-2}$	2.929216
Data 12: a^*	ARIMA (0,1,2)	$\hat{X}_t = X_{t-1} - 0.2508988e_{t-1} - 0.3525548e_{t-2}$	1.475189
Data 12: b^*	ARIMA (0,1,0)	$\hat{X}_t = X_{t-1}$	3.374610

According to the Box-Jenkins' method, model building starts by analyzing the autocorrelation function (ACF) and partial autocorrelation function (PACF) to identify the model. This is followed by parameter estimation and finally, model validation. However, this method is laborious, time-consuming, and expensive. Therefore, an alternative method is proposed [22].

3.3 LR Models

Since our data are positive, a simple and fast way to build a model for each data is by investigating whether the logarithmic returns are independent and normally distributed. To confirm the independency, the Durbin-Watson's test is used. At 5% significance level, Durbin-Watson's test confirmed that autocorrelation does not present, and the results are shown in Table 4.

Table 4. Durbin Watson test for twelve datasets.

Dataset	Sample size, n	D_L	D_U	Durbin Watson, D
Data 1	97	1.648510	1.690120	1.694643
Data 2	107	1.666000	1.703690	2.275297
Data 3	191	1.752620	1.773660	2.226176
Data 4	61	1.552400	1.618920	1.644443
Data 5	125	1.691910	1.724130	2.114983
Data 6	191	1.752620	1.773660	1.990135
Data 7	67	1.573780	1.634270	2.084813
Data 8	499	1.849140	1.857160	1.994098
Data 9	63	1.559870	1.624250	2.108231
Data 10	99	1.652230	1.692980	2.199957
Data 11	179	1.744190	1.766650	2.002313
Data 12: L^*	111	1.672310	1.708630	2.122952
Data 12: a^*	111	1.672310	1.708630	2.133196
Data 12: b^*	111	1.672310	1.708630	2.172809

While to confirm the normality, the Anderson-Darling's test is used. At 5% significance level, the results show that the normality assumption cannot be rejected.

Table 5. Anderson Darling test for twelve datasets.

Dataset	Anderson Darling, AD	p-value
Data 1	0.581582	0.126823
Data 2	0.604926	0.113290
Data 3	0.327174	0.516948
Data 4	0.416510	0.322015
Data 5	0.737939	0.053140
Data 6	0.614254	0.108611
Data 7	0.284300	0.619913
Data 8	0.286709	0.621444
Data 9	0.425040	0.307496
Data 10	0.295427	0.589104
Data 11	0.659116	0.083989
Data 12: L^*	0.308835	0.553281
Data 12: a^*	0.658723	0.083337
Data 12: b^*	0.270202	0.670823

From the data analysis, it is important to note that the behaviors of logarithmic return data are i.i.n.d. This is an indication that the original data can be adequately described by using the LR model. By using Eq. (1), we derive the LR model for each data and calculate its MAPE. The results are given in Table 6.

Table 6. LR model for twelve datasets.

Dataset	LR Model	MAPE (%)
Data 1	$\hat{X}_t = e^{-0.000033} \cdot X_{t-1} \cdot \left(\frac{X_{t-1}}{X_{t-2}}\right)^{0.132299}$	0.098489
Data 2	$\hat{X}_t = e^{-0.026332} \cdot X_{t-1} \cdot \left(\frac{X_{t-1}}{X_{t-2}}\right)^{-0.168256}$	23.343501
Data 3	$\hat{X}_t = e^{0.002583} \cdot X_{t-1} \cdot \left(\frac{X_{t-1}}{X_{t-2}}\right)^{-0.115495}$	13.767008
Data 4	$\hat{X}_t = e^{0.009425} \cdot X_{t-1} \cdot \left(\frac{X_{t-1}}{X_{t-2}}\right)^{0.160394}$	29.513728
Data 5	$\hat{X}_t = e^{0.015080} \cdot X_{t-1} \cdot \left(\frac{X_{t-1}}{X_{t-2}}\right)^{-0.102125}$	5.458634
Data 6	$\hat{X}_t = e^{0.001557} \cdot X_{t-1} \cdot \left(\frac{X_{t-1}}{X_{t-2}}\right)^{-0.002903}$	7.081969
Data 7	$\hat{X}_t = e^{0.000045} \cdot X_{t-1} \cdot \left(\frac{X_{t-1}}{X_{t-2}}\right)^{-0.065265}$	0.350150
Data 8	$\hat{X}_t = e^{0.000100} \cdot X_{t-1} \cdot \left(\frac{X_{t-1}}{X_{t-2}}\right)^{-0.003439}$	0.144578
Data 9	$\hat{X}_t = e^{0.064255} \cdot X_{t-1} \cdot \left(\frac{X_{t-1}}{X_{t-2}}\right)^{-0.165523}$	27.584105
Data 10	$\hat{X}_t = e^{-0.000027} \cdot X_{t-1} \cdot \left(\frac{X_{t-1}}{X_{t-2}}\right)^{-0.128282}$	1.005437
Data 11	$\hat{X}_t = e^{0.002892} \cdot X_{t-1} \cdot \left(\frac{X_{t-1}}{X_{t-2}}\right)^{-0.010728}$	5.228882
Data 12: L^*	$\hat{X}_t = e^{-0.001450} \cdot X_{t-1} \cdot \left(\frac{X_{t-1}}{X_{t-2}}\right)^{-0.065064}$	3.191593
Data 12: a^*	$\hat{X}_t = e^{-0.001004} \cdot X_{t-1} \cdot \left(\frac{X_{t-1}}{X_{t-2}}\right)^{-0.095076}$	1.585084
Data 12: b^*	$\hat{X}_t = e^{-0.000811} \cdot X_{t-1} \cdot \left(\frac{X_{t-1}}{X_{t-2}}\right)^{-0.093010}$	3.379359

The last column of Table 6 shows the prediction accuracy of each model. According to this table, most of the models are of high accuracy with MAPE values between 0.098489% and 7.081969%. Data 3 is accurate with MAPE 13.767008% while Data 2, Data 4 and Data 9 give reasonable accuracy with MAPE values between 23.343501% and 29.513728%.

3.4 Discussion

In this study, the LR model provides the desired model for each data. To support this claim, its performance was compared with the ARIMA model. In comparison between Tables 3 and 6, in terms of mathematical equations, it is apparent that the LR model was simpler than the ARIMA model. Based on the parsimonious principle, the former model is preferable than the latter. The finding shows that the LR model has two parameters involved. Meanwhile, the ARIMA model has up to six parameters in different data. In terms of prediction of accuracy, the LR model was comparably good with ARIMA model for all data. In terms of computation, LR modeling involves (i) logarithmic transformation to compute the logarithmic return data; and (ii) estimate the parameter of the simple linear regression in Eq. (1). On the contrary, to find the best ARIMA model, it involves an iterative process which might be time consuming and costly. Therefore, if time series data can be described as accurately as desired by employing LR model, then Box-Jenkins' methodology by using ARIMA model can be excluded; LR model is appropriate for model building with remarkably lesser effort and without compromising its desirable accuracy. Furthermore, a simple linear regression is sufficient to determine the parameters in the LR model. On the other hand, the ARIMA model needs a complicated computation process. This makes the running time of the proposed model shorter than the ARIMA model. In general, by using *R*-programming language, less than 0.20 seconds (in CPU time) were needed to obtain an LR model for each dataset. However, to construct an ARIMA model, more than 5 seconds (in CPU time) must be spent for every dataset. Thus, the LR model was not only simpler, comparably as good as to the ARIMA, but also faster than the ARIMA model.

4. CONCLUSIONS

With the GBM law as the basis, the LR model is derived to fit the autocorrelated data. Upon promising results obtained from fitting it to a cocoa powder dataset, the LR model is then applied to eleven more secondary real datasets from eight different fields and benchmarked with the ARIMA model using *R*-programming language. For all the datasets, the findings disclosed that the LR model is as good as ARIMA in terms of prediction accuracy; and performs much faster (less than 0.20 seconds for the LR model when compared to more than 5 seconds for the ARIMA model) in terms of CPU running time. Moreover, the LR model showcases a computational advantage over ARIMA model, particularly evident in larger sample sizes, owing to the iterative complexity involved in ARIMA modeling. On top of that, it is obvious that the LR model is an easy-to-compute and parsimonious methodology since it consists of two parameters and requires a two-step implementation procedure: logarithmic transformation to compute the logarithmic return data and parameter estimation of a simple linear regression. On the contrary, its counterpart has up to six parameters depending on the nature of the dataset and demands a time-consuming iterative implementation procedure involving model identification, parameter estimation, and model validation. This study shows that the LR model is a promising alternative methodology of ARIMA for positive datasets in time series modeling.

REFERENCES

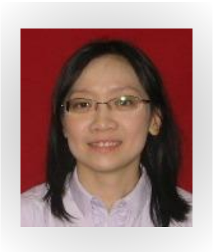
1. D. C. Montgomery, "Quality improvement in the modern business environment," in *Introduction to Statistical Quality Control*, 6th ed., John Wiley, NY, 2009, pp. 3-44.
2. Q. P. He and J. Wang, "Statistical process monitoring as a big data analytics tool for smart manufacturing," *Journal of Process Control*, Vol. 67, 2018, pp. 35-43.
3. B. Bouslah, A. Gharbi, and R. Pellerin, "Joint production, quality and maintenance control of a two-machine line subject to operation-dependent and quality-dependent failures," *International Journal of Production Economics*, Vol. 195, 2018, pp. 210-226.
4. E. M. Christino, G. M. Bonduelle, and S. Iwakiri, "Application of control charts in the production process of tauari (Couratari oblongifolia) wood flooring," *Cerne*, Vol. 16, 2010, pp. 299-304.
5. J. P. Guarnieri, A. M. Souza, L. F. Jacobi, B. Reichert, and C. P. Veiga, "Control chart based on residues: Is a good methodology to detect outliers?" *Journal of Industrial Engineering International*, Vol. 15, 2019, pp. 119-130.
6. C. Superville, "Outlier detection in autocorrelated manufacturing processes," *International Journal for Quality and Productivity Management*, Vol. 11, 2014, pp. 1-11.
7. W. H. Woodall and F. W. Faltin, "Autocorrelated data and SPC" *American Society for Quality Control (ASQC) Statistics Division Newsletter*, Vol. 13, 1993, pp. 18-21.
8. M. Kovářík and P. Klímek, "The usage of time series control charts for financial process analysis," *Journal of Competitiveness*, Vol. 4, 2012, pp. 29-45.
9. H. H. Ang, M. L. Huang, C. M. Lai, and J. R. Jin, "An approach combining data mining and control charts-based model for fault detection in wind turbines," *Renewable energy*, Vol. 115, 2018, pp. 808-816.
10. R. G. Aykroyd, V. Leiva, and F. Ruggeri, "Recent developments of control charts, identification of big data sources and future trends of current research," *Technological Forecasting and Social Change*, Vol. 144, 2019, pp. 221-232.
11. J. L. Vivancos, R. A. Buswell, P. Cosar-Jorda, and C. Aparicio-Fernandez, "The application of quality control charts for identifying changes in time-series home energy data," *Energy and Buildings*, Vol. 215, 2020, p. 109841.
12. I. Sanchez and I. Gonzalez, "Monitoring shrimp growth with control charts in aquaculture," *Aquacultural Engineering*, Vol. 95, 2021, p. 102180.
13. G. E. P. Box, G. M. Jenkins, and G. C. Reinsel, *Time Series Analysis: Forecasting and Control*, 4th ed., John Wiley and Sons, NY, 2008.
14. J. A. Fernandes de Souza, M. M. Silva, S. G. Rodrigues, and S. M. Santos, "A forecasting model based on ARIMA and artificial neural networks for end-OF-life vehicles," *Journal of Environmental Management*, Vol. 318, 2022, p. 115616.
15. D. Tomić and S. Stjepanović, "Forecasting capacity of ARIMA models: A study on Croatian industrial production and its sub-sectors," *Zagreb International Review of Economics and Business*, Vol. 20, 2017, pp. 81-99.
16. S. L. Lee, C. Y. Liew, C. K. Chen, and L. L. Voon, "Geometric Brownian motion-based time series modeling methodology for statistical autocorrelated process control: Logarithmic return model," *International Journal of Mathematics and Mathematical Sciences*, Vol. 2022, 2022, No. 4783090.
17. O. Awe, A. Okeyinka, and J. O. Fatokun, "An alternative algorithm for ARIMA model selection," in *Proceedings of IEEE International Conference in Mathematics, Com-*

puter Engineering and Computer Science, 2020, pp. 1-4.
18. Y. E. Shao and J. T. Dai, "Integrated feature selection of ARIMA with computational intelligence approaches for food crop price prediction," *Complexity*, Vol. 2018, 2018, No. 1910520.
19. D. Eni and F. J. Adeyeye, "Seasonal ARIMA modeling and forecasting of rainfall in Warri Town, Nigeria," *Journal of Geoscience and Environment Protection*, Vol. 3, 2015, pp. 91-98.
20. R. Adhikari and R. K. Agrawal, "Basic concepts of time series modeling," in *An Introductory Study on Time Series Modeling and Forecasting*," LAP Lambert Academic Publishing, Germany, 2013, pp. 12-17.
21. P. E. Omaku, O. J. Braimah, A. A. Adesupo, and S. B. Jaiyeola, "On the comparison of some models for estimating autocorrelated time series," *American Journal of Mathematics and Statistics*, Vol. 6, 2016, pp. 9-17.
22. S. L. Lee, C. Y. Liew, C. K. Chen, and L. L. Voon, "A modeling methodology for positive autocorrelated process data," in *Proceedings of IEEE International Conference on Computer and Drone Applications*, 2022, pp. 7-11.
23. K. D. Lawrence, R. K. Klimberg, and S. M. Lawrence, *Fundamentals of Forecasting Using Excel*, Industrial Press, NY, 2009.
24. M. B. Rao and C. R. Rao, *Computational Statistics with R*, 1st ed., Elsevier, 2014.
25. A. C. Harvey, *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press, NY, 1989.
26. P. J. Brockwell and R. A. Davis, *ITSM for Windows: A User's Guide to Time Series Modelling and Forecasting*, Springer-Verlag, NY, 1994.
27. J. D. Cryer and K. S. Chan, *Time Series Analysis: With Applications in R*, 2nd ed., Springer, NY, 2008.
28. D. C. Montgomery, C. L. Jennings, and M. Kulahci, *Introduction to Time Series Analysis and Forecasting*, Wiley-Interscience, NY, 2008.

Siaw Li Lee (李小麗) received her Ph.D. degree in Mathematics from Universiti Teknologi Malaysia in 2016. She is currently a Senior Lecturer at Universiti Teknologi MARA (UiTM), Sarawak Branch. Her research interests include time series modeling and multivariate statistical process control.

Chin Ying Liew (劉靜盈) received her Ph.D. degree in Computational Science from Universiti Malaysia Sarawak in 2016. She is currently a senior Lecturer at Universiti Teknologi MARA (UiTM), Sarawak Branch. Her research interests focus on computational modeling generally, bipartite network modeling specifically, and meaningful learning experience of students across education levels, particularly in mathematics.

Chee Khium Chen (曾志瓊) received his Master of Science in Land Administration and Development, and Bachelor of Engineering (Geomatics) from Universiti Teknologi Malaysia. He is currently a Senior Lecturer at Universiti Teknologi MARA (UiTM), Sarawak Branch. His research interests include mathematics, statistics, and geomatics engineering.



Li Li Voon (温麗麗) received her MSc. in Mathematics from Universiti Sains Malaysia. She is currently a senior lecturer at Universiti Teknologi MARA (UiTM), Sarawak Branch. Her research interests are on the areas of mathematics teaching and learning, and mathematics innovation.