CONTROL CHART AS VERIFICATION TOOLS IN TIME SERIES MODEL

Nurfitri Imro'ah [1], Nur'ainul Miftahul Huda [2*]

[1]Statistics Department, FMIPA, Universitas Tanjungpura
Prof. Dr. H. Hadari Nawawi St., Pontianak City, West Kalimantan, 78124, Indonesia
[2]Mathematics Department, FMIPA, Universitas Tanjungpura
Prof. Dr. H. Hadari Nawawi St., Pontianak City, West Kalimantan, 78124, Indonesia

Corresponding author's e-mail: [2*] nur'ainul@fmipa.untan.ac.id

Abstract. *Control charts are generally use in quality control processes, especially in the industrial sector, because they are helpful to increase productivity. However, control charts can also be used in time series analysis. The residuals from the time series model are used as observations in constructing the control chart. Because there is only one variable observed, namely the residual, the control chart used is the Individual Moving Range (IMR). This study analysis the accuracy of the time series model using the IMR control chart in two models, namely the Autoregressive Distributed Lag (ADL) model without outliers and the ADL model with outliers. The results showed that the control chart could be used to measure the accuracy of the time series model. The accuracy of the model can be seen from the statistically controlled residual (in control).*

Keywords: *Individual Moving Range, Autoregressive Distributed Lag, residual.*

https://ojs3.unpatti.ac.id/index.php/barekeng/   barekeng.math@yahoo.com

1. INTRODUCTION

The control chart in Statistical Process Control (SPC) is used to determine whether the process is in control or not [1]. Two types of control charts are variable and attribute control charts. Some researchers use control charts as a tool to control the production process. Haddad (2021) reviews the ability of the Palestinian construction companies to apply the essential tools of SPC in their concrete production systems for process improvement [2]. Hsu, Wang, Lin, Chen and Hsu (2020) applies statistical process control and machine learning techniques to diagnose wind turbine faults and predict maintenance [3]. However, the application of control charts can also be carried out in time series analysis uses here Box-Jenkins iterative steps: order identification, parameter estimation, and diagnostic test [4]. The best model in the time series analysis is obtained in the diagnostic test step, where the model must fulfill the white noise assumption, namely the independent and normality of the residuals [5].

Measuring the model's accuracy on the best model is very important in predicting the following several periods. The accuracy of a model is seen from the residuals is value [6]. The smaller the residual, the more accurate the model. Alternatively, it can be said that the residual is in a statistically controlled state. Such a situation can be seen in the control chart. So the role of the control chart in the time series analysis is in measuring the accuracy of the model obtained.

The case study in this paper is based on [7], which models the number of dengue fever cases in West Kalimantan using the Autoregressive Distributed Lag (ADL) model with outlier factors. Data about the number of dengue fever cases were first analyzed using the Box-Jenkins method to obtain an ADL model without outliers. Hereafter, outlier detection is done then the ADL model with outlier factors is obtained. The selection of the best model in this study was based on a model with all significant parameters.

This paper aims to verify the best model obtained in [7] using the control chart based on the residuals from the model. Two models were verified in this study, namely the ADL model without outlier factors and the ADL model with outlier factors. The first section describes the background. An overview of control charts and time series control charts is portrayed sequently in the second and third sections. In comparison, the fourth section explains the case studies and data analysis. While the conclusions are put forward in the fifth section.

2. RESEARCH METHOD

2.1   Control Chart

The control chart is a technique used to evaluate a process, whether the process is statistically in control or not, to solve problems and produce quality improvements. Three lines in control chart, namely Center Line (CL), Upper Control Limit (UCL) and Lower Control Limit (LCL) [8]. Center Line (CL) is the average value of quality characteristics. While UCL and LCL are the control limit. Fig. 1 is the examples of a control chart for some conditions. The process is said to be out of control if [9]:
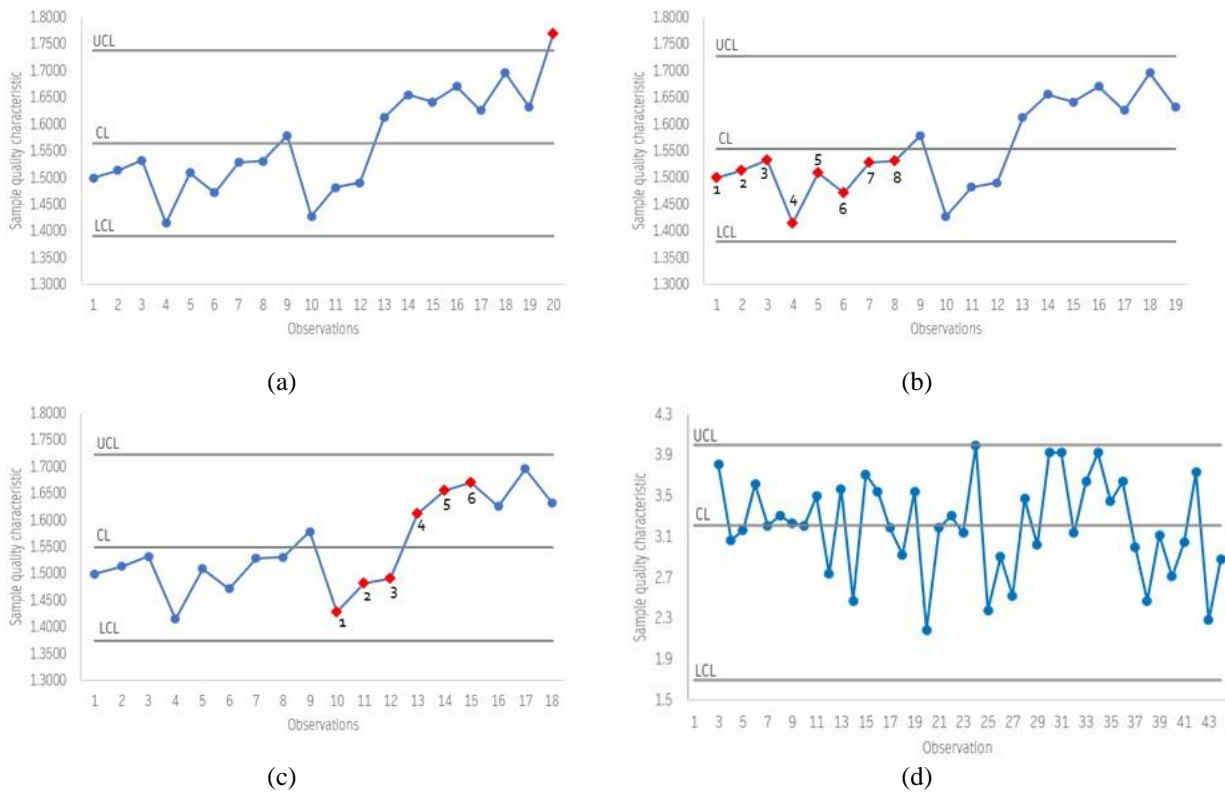
1) There is a point that is beyond the control limits (see Fig. 1a).
2) At least seven consecutive points located above or below CL (see Fig. 1b).
3) There are six or seven points in a row that keep going up or down (see Fig. 1c).

If the process is out of control, it is necessary to investigate and analyze the cause. While if the sample points are within the control limits, then the process is in control (see Fig. 1d) [10]. Let w be a statistical sample that measures some quality characteristic and let μ_w be the mean of w and the standard deviation of w denoted σ_w. Then UCL, CL and LCL are expressed as follows [11].

$$UCL = \mu_w + k\sigma_w$$
$$CL = \mu_w$$
$$LCL = \mu_w - k\sigma_w$$

where k is the distance of the control limit from the CL, expressed in standard deviation units. Walter A. Shewhart firstly introduced the general theory of the control chart, accordingly the control chart based on these principles is called the Shewhart control chart. Control charts are classified into two types, namely

variable and attribute control charts [12]. The difference between the two types lies in whether the characteristics are measurable (for variable control chart) or not (for attribute control chart). Three types variable control charts, namely $\bar{X} - R$ (for small sample size or < 10), $\bar{X} - s$ (for large sample size or > 10) and IMR (for single sample size) [13].

(a)

(b)

(c)

(d)

Figure 1. The examples of a control chart showing an out of control process caused by (a) There is a point that is beyond the control limits, (b) There are eight consecutive points located below CL, (c) There are six points in a row that keep going up. While the process in control is shown in (d).

2.2 Time Series Control Chart

The basic assumptions on the control chart are independence and no correlation among the observations [14]. In the time series control chart the observations used are residuals (due to the residuals assumption in the time series model is white noise, which is independent and uncorrelated among residuals), considering that only one variable is observed that is the residual model. Consequently, constructing a control chart for the time series model can be applied to any model that assumes white noise on the residuals. The control chart that fulfill these condition is the IMR control chart, which the calculations can be written as [15].

$$Individual\ plot = \begin{cases} UCL = \bar{x} + 3\dfrac{\overline{MR}}{d_2} \\ CL = 0 \\ LCL = \bar{x} - 3\dfrac{\overline{MR}}{d_2} \end{cases} \tag{1}$$

and

$$Movingrange\ plot = \begin{cases} UCL = D_4 \times \overline{MR} \\ CL = \overline{MR} \\ LCL = 0 \end{cases} \tag{2}$$

where $\overline{MR}$ is the average of MR_i with $MR_i = |x_i - x_{i-1}|$, x_i is i-th observation, d_2 and D_4 are constants.

The residual (e_t), independent and identical normal distributed with zero mean and σ_e^2 is used in the time series control chart then the $\bar{x}$ in Eq. (1) equals to zero. The following are the steps to construct a time series control chart.

1) Calculate the moving range (MR_i) and average moving range $\overline{MR}$.
2) Compute the UCL, CL and LCL (see Eq. (1) and (2)).
3) Construct an IMR control chart consisting of individual plots and moving range plots. Then, add the UCL, CL and LCL calculated in second step in each plot. Residuals are contained in the individual plots, while the moving range plots contain the moving ranges calculated in the first step.
4) If the residuals are out of control, then the model obtained has poor accuracy. It is necessary to evaluate the identification of the time series model. If the residuals are in control, then the time series model obtained can be used for prediction, or it can be said that the model has good accuracy.

To overcome the residuals that are out of control, the time series model needs to be evaluated (in this paper, the model is evaluated by adding the outlier factor).

For more details, the procedures in constructing a time series control chart can be seen in Fig. 2b. Fig. 2a describes the outlier detection algorithm in the ADL model. After the outlier detection time is obtained, the outlier factor is added to the ADL model. The residual of the ADL model with an outlier factor is the input to flowchart 2b. The purpose of the evaluation step in Fig. 2b is that if the residuals are out of control, then the step returns to Fig. 2a and is adjusted to the model used.
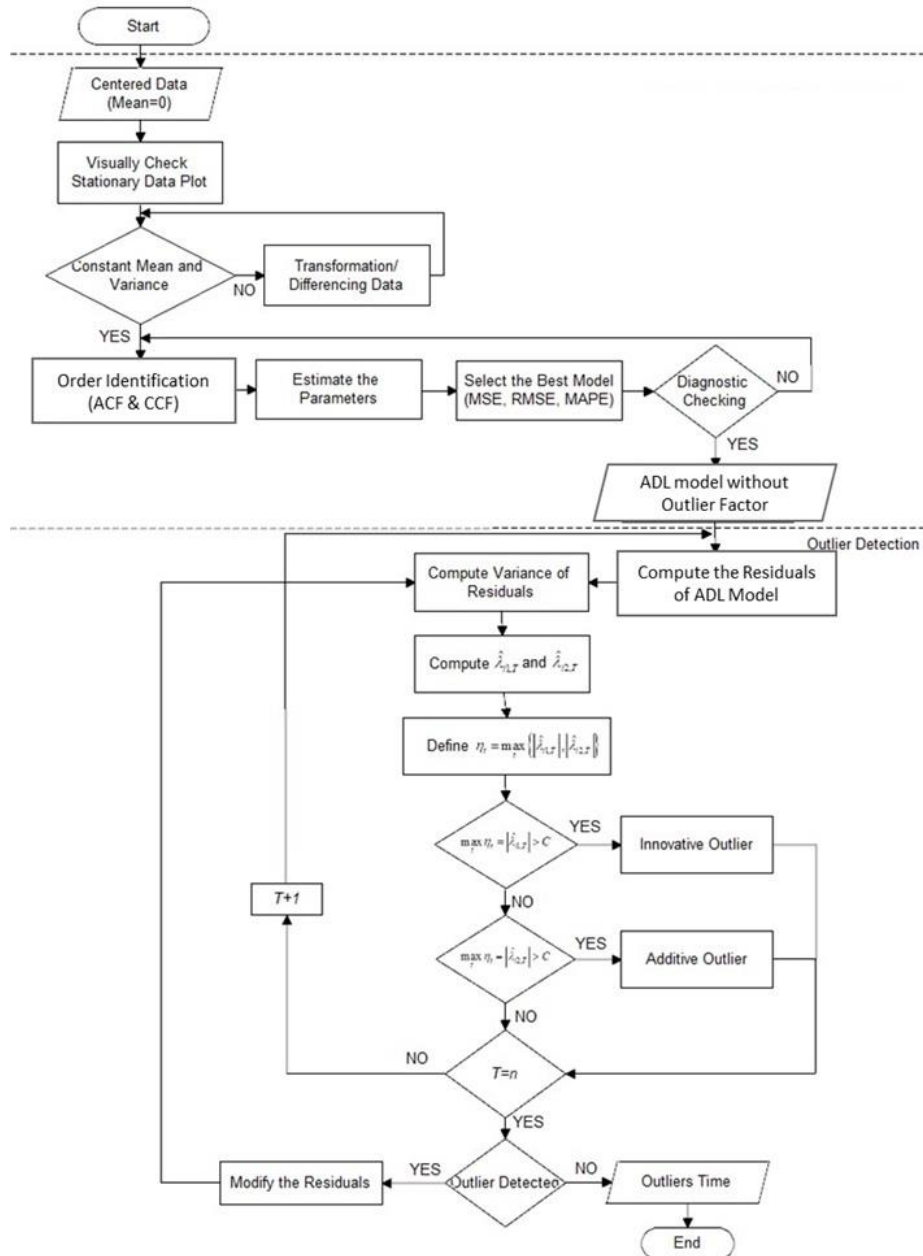

(a)
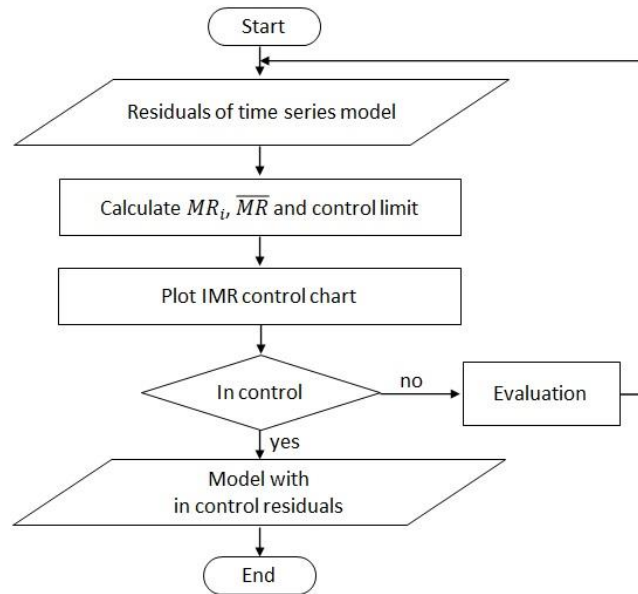
(b)

Figure 2. Flowchart, (a) Flowchart of ADL model with outlier factor [7]**, and (b) Flowchart of control chart for time series model**

3. RESULT AND DISCUSSION

This study uses a case study in [7], which uses the ADL model to model the number of dengue fever cases in West Kalimantan. Two models were obtained, namely the ADL model without outlier factors (referred to Model 1) and the ADL model with outlier factors (referred to Model 2). The Box Jenkins method was applied to the data to obtain Model 1, while Model 2 was obtained based on detecting outliers in Model 1. In this study, the residuals from each model were used to construct the IMR control chart. Fig. 3 is an IMR control chart for Model 1.

(a)   (b)
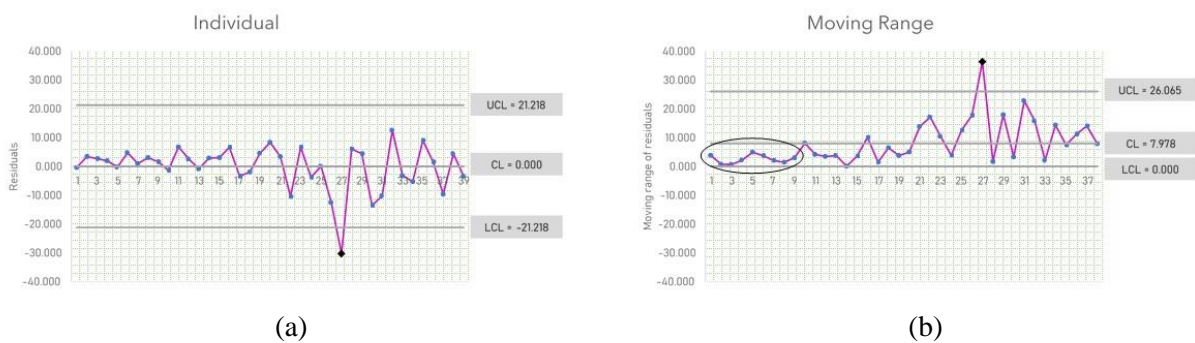
Figure 3. IMR Control Chart for Model 1, (a) Individual plot and (b) Moving range plot

Fig. 3 shows that the residuals Model 1 are out of control because there is one point outside the control limit (see Fig. 3a and 3b), and there are nine points in a row between CL and LCL (see Fig. 3b). Residuals that are outside the control limits indicate that the residual value is much different from other residuals. So that, the difference between the observations and the estimation results become large. One of the reasons for this is the presence of outliers in the data. Therefore Model 1 is evaluated by adding an outlier factor. The ADL model with outlier factors is referred to as Model 2. Residuals Model 2 are then remade the control chart to determine whether the residuals are in control or out of control. The control chart for Model 2 is presented in Fig. 4 below.
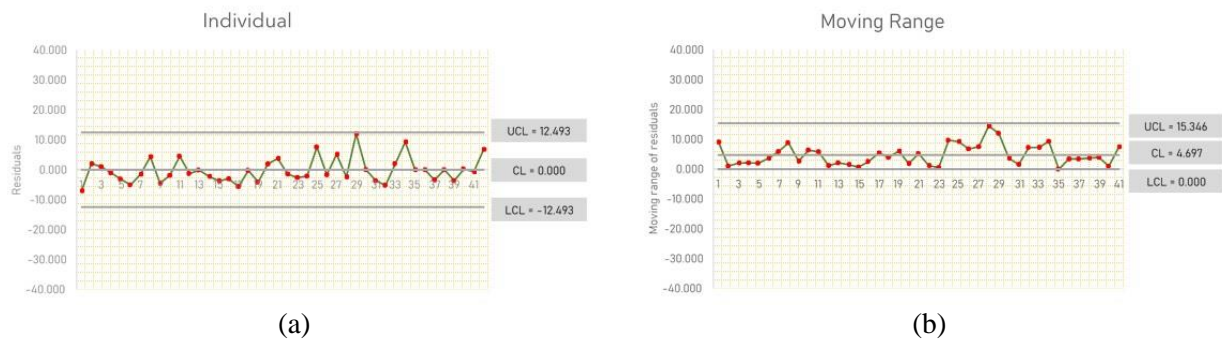
Figure 4. IMR Control Chart for Model 2, (a) Individual plot and (b) Moving range plot

Based on Fig. 4, there is no point that is outside the control limit, or it can be said that the residuals are in control. So, it can be concluded that Model 2 is an accurate model for predicting several future periods. Alternatively, it can be said that Model 2 is the best model because the residuals are in a statistically controlled state.

So that consistent with the conclusion in [7] that the presence of outliers makes the model less accurate (resulting in out of control residuals). So that the addition of an outlier factor into the model is one of the proper steps to produce a more accurate model (marked by residuals in control); thus, the control chart can be used as a tool to measure the model's accuracy from the residuals perspective.

4. CONCLUSIONS

There are two things obtained based on the analysis that has been done. First, the presence of outliers in the data causes the residual value to be too large or too small compared to other residuals so that the model's accuracy is low. It can be seen from the residual model on the control chart that it is not statistically controlled (out of control). Second, the addition of outliers to the model has a good impact on the model's accuracy. The model obtained becomes more accurate because the residuals of the model are in control. From these two things, it can be concluded that the control chart can be used as a tool to analyze the accuracy of the time series model. The appropriate control chart is IMR, with the residuals model as observations on the control chart.

AKNOWLEDGEMENT

REFERENCES

[1]   D. C. Montgomery, *Introduction to Statistical Quality Control*, 5th ed. New York: John Wiley and Sons, 2005.

[2]   T. Haddad, "Quality Assessment of Concrete Production Using Statistical Process Control (Spc) Techniques," *Proc. Eng. Sci.*, vol. 3, no. 2, pp. 239–246, 2022, doi: 10.24874/pes03.02.011.

[3]   J. Y. Hsu, Y. F. Wang, K. C. Lin, M. Y. Chen, and J. H. Y. Hsu, "Wind turbine fault diagnosis and predictive maintenance through statistical process control and machine learning," *IEEE Access*, vol. 8, pp. 23427–23439, 2020, doi: 10.1109/ACCESS.2020.2968615.

[4]   G. E. P. Box and G. M. Jenkins, "Time series analysis; forecasting and control [by] George E. P. Box and Gwilym M. Jenkins." pp. xix, 553 p., 1970.

[5]   J. D. Cryer and K. S. Chan, *Time Series Analysis: With Application in R*, 2nd ed. USA: Springer Science and Business Media, 2008.

[6]   W. W. . Wei, *Time Series Analysis: Univariate and Multivariate Methods*, 2nd ed. New Jersey: Pearson Prentice Hall, 2006.

[7]   N. M. Huda, U. Mukhaiyar, and U. S. Pasaribu, "Forecasting dengue fever cases using autoregressive distributed lag model with outlier factor," *AIP Conf. Proc.*, vol. 2268, no. September, 2020, doi: 10.1063/5.0018450.

[8]   F. M. Vilas Boas *et al.*, "Condition Monitoring of Internal Combustion Engines in Thermal Power Plants Based on Control Charts and Adapted Nelson Rules," *Energies*, vol. 14, no. 16, p. 4924, Aug. 2021, doi: 10.3390/en14164924.

[9]   M. Inkelas *et al.*, "Using control charts to understand community variation in COVID-19," *PLoS One*, vol. 16, no. 4, p. e0248500, Apr. 2021, doi: 10.1371/journal.pone.0248500.

[10]   H. Li, J. Deng, S. Yuan, P. Feng, and D. D. K. Arachchige, "Monitoring and Identifying Wind Turbine Generator Bearing Faults Using Deep Belief Network and EWMA Control Charts," *Front. Energy Res.*, vol. 9, Nov. 2021, doi: 10.3389/fenrg.2021.799039.

[11]   J. . Oakland, *Statistical Process Control*. Butterworth-Heinemann, 2003.

[12]   S. Chakraborti and M. A. Graham, "Nonparametric (distribution-free) control charts: An updated overview and some results," *Qual. Eng.*, vol. 31, no. 4, pp. 523–544, Oct. 2019, doi: 10.1080/08982112.2018.1549330.

[13]   S. Ali, Z. Abbas, H. Z. Nazir, M. Riaz, X. Zhang, and Y. Li, "On developing sensitive nonparametric mixed control charts with application to manufacturing industry," *Qual. Reliab. Eng. Int.*, vol. 37, no. 6, pp. 2699–2723, Oct. 2021, doi: 10.1002/qre.2885.

[14]   B. Elahi, A. Deheshvar, M. Franchetti, and S. Amiri Tokaldany, "Optimum evaluation performance of adaptive X-bar control charts," *J. Ind. Prod. Eng.*, pp. 1–10, Jun. 2021, doi: 10.1080/21681015.2021.1944341.

[15]   L. C. Alwan and H. V. Roberts, "Time-series modeling for statistical process control," *J. Bus. Econ. Stat.*, vol. 6, no. 1, pp. 87–95, 1988, doi: 10.1080/07350015.1988.10509640.