DOI: 10.56892/bima.v8i1.634

Arithmetic Return model for removing Autocorrelation from Statistical process control data exhibiting Geometric Brownian Motion

Chajire B. P.[1]*, I. J. Dike[2], S. S. Abdulkadir[2] and E. Torsen[2]

[1]Department of Mathematics, Faculty of Science, Gombe State University, PMB 127 Gombe. Nigeria

[2]Department of Statistics, Faculty of Physical Sciences, Modibbo Adama University Yola,Nigeria

Corresponding Aurhor email:

ABSTRACT

The presence of positive autocorrelation in a controlled process is a major problem especially when traditional quality control charts are to be used in monitoring the process. This is because the two major assumptions in using the traditional control charts are that the process data are independently and normally distributed. In this work, a novel method of removing autocorrelation from data exhibiting Geometric Brownian Motion (GBM) is proposed. This GBM is autoregressive of order AR(1). A chemical process dataset and furnace temperature dataset were transformed to Arithmetic Return model (ARM). The fitted ARM for both datasets were fitted and residuals obtained from both datasets were subjected to DW test for the presence of positive autocorrelation. Initial Durbin Watson's (DW) test result for both processes before the transformation were 0.0538 and 1.5045 respectively which indicated the presence of positive autocorrelation. Final DW test results from the ARM transformation were 2.0047 and 1.7848 respectively indicating that positive autocorrelation was removed from both datasets. The proposed method is simple to understand and easy to use provided that the process data is GBM and autocorrelation is the major concern.

Keywords: Autocorrelation, Geometric Brownian Motion (GBM), Autoregressive of order 1 AR(1), Arithmetic Return Model (ARM), Logarithmic Return Model (LRM), Durbin Watson (DW), Statistical Process Control.

INTRODUCTION

Statistical quality control charts have been designed to work with the two basic assumptions that the process data are normally distributed and that they should be independent. In essence, a process outcome data should not determine the outcome of the next successive process data and so on. This is established in the literature but in real life, some process that are being controlled are autocorrelated in nature. Most of such processes are time dependent.

Autocorrelation is a representation of the degree of similarity mathematically, between a given time series data and a lagged version of itself over successive time intervals (Tim, 2023). It is conceptually similar to the correlation between two different time series, but autocorrelation uses the same time series twice. Once in its original form and lagged in one or more time periods.

A process is only said to be in a state of statistical control if it is operating in the presence of natural causes of variation and is free from assignable causes (Dike *et al.,* 2018). When autocorrelation is present in the process, unnecessary out of control signals may be triggered which makes a problem that has to be treated with urgency before monitoring of the process.

In the continuous industries such as the chemical industry, most process data are autocorrelated. In other circumstances, such as when a process follows an adaptive model or when the process mean is a deterministic function, for example, a harmonic function or a linear or nonlinear trend, the data will also be autocorrelated (Nien, 1999).

Douglas (2013) stated that the most important of the assumptions made concerning control charts is that of independence of the observations, for conventional control charts do not work well if the quality characteristic exhibits even low levels of correlation over time. Specifically, these control charts will give misleading results in the form of too many false alarms if the data are positively correlated. To avoid such unnecessary disturbances, the process data needs to be cleared of autocorrelation before monitoring.

In discrete as well as in continuous production processes, data often show some autocorrelation, or serial dependence. Several monitoring tools are found in the literature that deals with the case of multiple process variables, but a few of them deal with the case of autocorrelated data (Hussam *et al.,* 2021).

GEOMETRIC BROWNIAN MOTION (GBM)

A process X is said to follow a Geometric Brownian Motion, $= \{X_t: t \in [0, \infty)\}$, with a constant volatility σ and constant drift μ if it satisfies the stochastic differential equation,

$$dX_t = \sigma X_t dW_t + \mu X_t dt$$

where W_t is a wiener process as given by Rathnayaka *et al.* (2014).

The GBM is a continuous time stochastic process in which the log of the random varying quantity follows a Brownian process (also called Wiener's process). $X_0 = 1$, so, the process starts at 1, but this can easily be changed. GBM is an example of a stochastic process which is used to model processes that can never take on negative values, such as the value of stocks, currency exchange rates and other financial instruments. This GBM is a mathematical model used to describe the random fluctuations in the price of financial assets over time (Kyle, 2022).

From literatures, the two major ways to remove autocorrelation from the process is by fitting a time series model to the process data, thereby obtaining independent residuals, and the residuals are now used as the process control data. Classica l control charts for independent variables can be applied to residual processes, provided that the residuals are independent Erna *et al.* (2018). Or, the traditional control charts limits are simply adjusted to account for autocorrelation in the process. Thirdly, a less popular approach is the model-free charts as stated by (Aytaç, 2020). These are control charts for autocorrelation that do not apply any modeling techniques to the data. They can be considered as a separate category of charts where no residuals are used and no modifications are done to existing charts.

The first method involved the use of some sophisticated statistical skills to build a satisfactory model which led to the work of Siaw *et al.* (2013) where a new method of fitting a time series model to the process with ease was proposed by using Logarithmic Return transformation. Djauhari *et al.* (2014) established that the positive time series data are governed by Geometric Brownian Motion law, and this GBM is an autoregressive process of order 1 AR(1). The derivations are found in Siaw *et al.* (2022).

This study is an exploration of the work done by Siaw *et al.* (2022), where Logarithmic return transformation was done on the AR(1) process to obtain i.i.n.d residuals. According to literatures, when the rate of return is not large, logarithmic return is approximately equal to the arithmetic return. This work is intended by using arithmetic return transformation instead of the logarithmic return on a GBM process to see the response.

The Proposed Arithmetic Return Model (ARM)

The proposed method is obtained by converting the AR (1) process to Arithmetic Return Model, this because the Geometric Brownian Motion (GBM) is an AR(1) process.

To consider the Arithmetic Return Model for dependent data, transformation of Arithmetic Return (AR) is needed and shown as

$$AR = \frac{X_t - X_{t-1}}{X_{t-1}} \qquad (1)$$

Consider the AR (1) process which can be written as;

$$AR_t = C + \Theta AR_{t-1} + \mathcal{E}_t \qquad (2)$$

where C is the intercept, Θ the slope in AR(1) model and the error terms $\mathcal{E}_t$ are i.i.n.d. with zero mean and constant variance.

Then the Arithmetic Return transformation of Equation (2) is given as

$$\frac{X_t - X_{t-1}}{X_{t-1}} = C + \Theta\left(\frac{X_{t-1} - X_{t-2}}{X_{t-2}}\right) + \mathcal{E}_t \qquad (3)$$

thus,

$$\left.\begin{aligned} \frac{X_t}{X_{t-1}} - 1 &= C + \Theta\left(\frac{X_{t-1}}{X_{t-2}} - 1\right) + \mathcal{E}_t \\ \frac{X_t}{X_{t-1}} &= C + 1 + \Theta\left(\frac{X_{t-1}}{X_{t-2}}\right) - \Theta + \mathcal{E}_t \\ \frac{X_t}{X_{t-1}} &= [C + 1 - \Theta] + \Theta\left(\frac{X_{t-1}}{X_{t-2}}\right) + \mathcal{E}_t \\ \frac{X_t}{X_{t-1}} &= \mathcal{B} + \Theta\left(\frac{X_{t-1}}{X_{t-2}}\right) + \mathcal{E}_t \end{aligned}\right\} \qquad (4)$$

where $\mathcal{B} = [C + 1 - \Theta]$, and Θ are the regression parameters, $\mathcal{B}$ is the intercept and Θ is the slope in the model.

Equation (4) is the Arithmetic Return model (ARM).

The residuals are obtained using Equation (5)

$$\widehat{\mathcal{E}}_t = X_t - \widehat{X}_t \qquad (5)$$

The Durbin Watsons test for autocorrelation

The Durbin-Watson (DW) test is a statistical test used to detect autocorrelation in the residuals of a linear regression model.

The null and alternative hypotheses are;

H_0 : There is no positive autocorrelation among the residuals.

H_1 : The residuals are autocorrelated

Test statistic:

$$d = \frac{\sum_{t=2}^{T} (e_t - e_{t-1})^2}{\sum_{t=1}^{T} e_t^2} \qquad (6)$$

where **T:** The total number of observations

e_t: The t[th] residual from the regression model

Decision rule for the Durbin Watsons (DW) test for the presence of positive autocorrelation

At significance level α (alpha), the test statistic DW is compared at lower and upper critical values.

If DW > dU, there is no statistical evidence that positive autocorrelation exists in the data

If DW < dL, there is statistical evidence that positive autocorrelation exists in the data.

If DW is between the two bounds, the test is inconclusive (CFI, 2024).

RESULTS

Data used for the analysis were obtained from simulations of a chemical process and a real-life process data used by (Salleh *et al.*, 2018). All analysis were carried out using Microsoft Excel 2016.

Results of Analysis from the Chemical Process Data

Table 1: Summary of results from the chemical process data

Statistic	Chemical process data	ARM	LRM
Slope	0.025249611	0.072793689	-0.00495605
Intercept	0.983811403	-0.00382924	-0.16939806
R squared	0.99977547	1.46633E-05	0.028703173
DW	0.053810486	2.003625015	2.338180758

$$\left(\frac{X_t}{X_{t-1}}\right) =- 0.00382924 + 0.072793689\left(\frac{X_{t-1}}{X_{t-2}}\right) \quad (7)$$

Equation (7) is the fitted ARM for the chemical process data. Estimated values for the regression parameters $\mathcal{B}$ and Θ are 0.072793689 and -0.00382924 respectively. From the Durbin Watson's table at 5% significance level, n = 200, k = 1, dL = 1.758 and dU = 1.779. Initial DW test was carried on the residuals of the process giving the value 0.053810486 which is less than dL indicating that positive autocorrelation exists in the process. ARM was fitted to the process dataset, and DW test was carried on the residuals obtained from the ARM which gave the value 2.003625015. The DW 2.003625015 > 1.779 which indicates that positive autocorrelation was removed from the chemical process data.

Results of Analysis from the Furnace Temperature Data

Table 2: Summary of results from the furnace temperature data

Statistic	Furnace data	ARM	LRM
Slope	0.723881222	0.145460909	0.14545076
Intercept	435.9849614	-4.8315×10^{-6}	-4.8926×10^{-6}
R squared	0.500212167	0.021732047	0.021729192
Correlation	0.70725679	0.147417933	0.14740825
DW	1.504463924	1.784796405	2.004645495

$$\left(\frac{X_t}{X_{t-1}}\right) =- 4.8315 \times 10^{-6} + 0.145460909\left(\frac{X_{t-1}}{X_{t-2}}\right) \quad (8)$$

Equation (8) is the fitted ARM for the furnace temperature data. Regression estimated values for $\mathcal{B}$ and Θ are -4.8315×10^{-6} and 0.145460909 respectively. From the Durbin Watson's table at 5% significance level, n = 78, k = 1 the lower and upper values are dL = 1.611 and dU = 1.662. Initial DW test value obtained from the residuals of the process data is 1.504463924 which is less than the dL indication that positive autocorrelation exists in the process data. ARM was fitted to the process dataset, and DW test was carried on the residuals obtained from the ARM which gave the value 1.784796405. The DW 1.784796405 > 1.662 which indicates that positive autocorrelation was removed from the chemical process data.

DISCUSSION

The initial DW test results from both datasets were 0.053810486 and 1.504463924 which indicated the presence of positive autocorrelation because both results fell below the DW Lower bounds.

After the ARM transformation of both datasets, DW test was carried again on the residuals that were obtained. Autocorrelation-free results of 2.003625015 and 1.784796405 which were confirmed to be above the DW upper bounds were obtained.

From results of the analysis, ARM proved effective in removing the effect of autocorrelation from both datasets used which were GBM.

CONCLUSION

In conclusion, the ARM has proven to be more effective in removing autocorrelation from both process data which are Geometric Brownian Motion with much simplicity in understanding and implementation than the LRM which has to go through the log return transformation of the process data. Further research is currently ongoing to find out other strengths of the ARM in statistical process control for autocorrelated data which is GBM driven.

REFERENCES

Aytaç, K. (2020). Control Charts for monitoring processes with Autocorrelated data. *Master's Dissertation submitted to obtain the degree of, Master in Business Engineering: Data Analytics*. Universitiet Gent.

CFI (2024). Durbin Watson Statistic. *Corporate Finance Institute Resources.* Retrieved on 29th February, 2024 from https://corporatefinanceinstitute.com/resources/data-science/durbin-watson-statistic/

Dike, I. J., Buba, C., and Dike, C.O. (2018). Models and Quality Control Charts for the Prediction of Compressive Cement Strength. *International Journal of Mathematics and Statistics Studies.* 6(2), 1-12. (www.eajournals.org)

Djauhari, M.A., Lee, S.L., and Ismail, Z. (2014). Model building for Autocorrelated process control: An industrial experience. *American Journal of Applied Sciences* 11 (6), 888-898. ISSN: 1546-9239 doi:10.3844/ajassp.2014.888.898

Douglas, C.M. (2013). Introduction to Statistical Quality Control (seventh edition). United Kingdom: John Wiley and Sons Inc.

Erna, T. H., Geysa, F., and Nurtiti, S. (2018). Modified Exponential Weighted Moving Average (EWMA) Control Chart on Autocorrelation Data. *IOP Conference Series: Journal of Physics*. 979(012097), 1-7. doi:10.1088/1742-6596/979/1/012097

Hussam, A., Mu'men, R., Tarek, A., and Omar., B. (2021). A Guassian process approach for monitoring autocorrelated batch production processes. *WILEY.* DOI: 10.1002/qre.2951

Kyle, S. (2022). Geometric Brownian Motion. Librettists Statistics. Retrieved on 15th November, 2023 from https://stats.libretexts.org

Nien, F. Z. (1999). Statistical control for Autocorrelated data. *Part of the*

EUROPTO Conference on Process and Equipment Control in Microelectronic Manufacturing. 3742, 0277-786.

Rathnayaka, R. M. K. T., Jianguo, W., and Seneviratna, D. M. K. N. (2014). Geometric Brownian Motion with Ito's Lemma Approach to Evaluate Market Fluctuations: A case study on Colombo Stock Exchange. *International Conference on Behavioral, Economic, and Socio-Cultural Computing (BESC2014).* doi:10.1109/besc.2014.7059517

Siaw, L. L., and Djauhari, M. A. (2013). Monitoring Autocorrelated Processes: A Geometric Brownian motion process approach. https://www.researchgate.net/publication/260829536 DOI: 10.1063/1.4823976

Siaw, L. L., Chin, Y. L., Chee, K. C., and Li. L. V. (2022). Geometric Brownian Motion-Based Time Series Modeling Methodology for Statistical Autocorrelated Process Control: Logarithmic Return Model. *International journal of Mathematics and Mathematical Sciences.* https://doi.org/10.1155/2022/4783090

Tim, S. (2023). Autocorrelation: What it is, how it works, Tests. Retrieved on 14th August 2023 from https://www.investopedia.com/terms/a/autocorrelation.asp.