



STATISTICAL PROCESS CONTROL OF AUTOCORRELATED DATA EXHIBITING GEOMETRIC BROWNIAN MOTION: ARITHMETIC RETURN MODEL APPROACH

*1Buba, C. P., ²Dike, I. J., ²Abdulkadir, S. S. and ²Torsen, E.

¹Department of Mathematical Sciences, Faculty of Science, Gombe State University ²Department of Statistics, Faculty of Physical Sciences, Modibbo Adama University Yola

*Corresponding authors' email: chajireb@gmail.com

ABSTRACT

Statistical process control charts operate with the two basic assumptions that the process data at different time points should be independent and identically normally distributed (i.i.n.d.). The assumption of independence is mostly violated which leads to wrong inferences made on the processes being monitored. This work aims at carrying out Statistical Process Control (SPC) of autocorrelated data exhibiting Geometric Brownian Motion (GBM): Arithmetic Return Model (ARM) approach. The proposed model proffers a simplified solution to the autocorrelation problem in SPC by transforming an Autoregressive of order 1 (AR (1)) process to ARM, this is because the GBM is the potential law which governs most positive time series, and has been confirmed to be AR(1) according to literature. Two sets of autocorrelated data which exhibit GBM properties; a simulated data and a furnace temperature data obtained from literature were subjected to SPC and monitoring. Findings from the work showed that the ARM performed better than the existing Logarithmic Return Model (LRM) in terms removal of autocorrelation faster with less sophistication, from 1.171261 to 2.86818 in the simulated data, and from 1.50446 to 1.7848 in the furnace temperature data. Also, the proposed model gave a better fitting of both process data, and faster detection ability for out of control signals in the control charts at λ =0.25, λ =0.7 which makes the proposed ARM a better choice when dealing with SPC of autocorrelated data exhibiting the GBM.

Keywords: Statistical process control, Geometric Brownian Motion (GBM), Autocorrelated data

INTRODUCTION

Control charts for process monitoring have traditionally been designed and evaluated under the assumption that observations on the process output at different times are independent (Chao-Wen & Marion, 2001). The most important of the assumptions made concerning control charts is that of independence of the observations (Montgomery, 2013). Control charts are the most popular monitoring tools used to monitor changes in a process and distinguished between assignable and chance causes of variations (Rupali & Vikas, 2022). When the assumption of independence is satisfied, conventional control charts may be applied to the process for monitoring and evaluation. However, the independence assumption is often violated in practice because the variables tend to have a reasonable level of correlation among them and this can produce a major impact on the process estimates thereby leading to faulty inferences about the process. The conventional control charts do not work well if the quality characteristic exhibits even low levels of correlation over time (Montgomery, 2013). Traditional control charts may produce misleading results, such as an increased frequency of false alarms, when applied to positively autocorrelated data.

Autocorrelation, also known as serial correlation, refers to the degree of correlation of the same variables between two successive time intervals. It is known for measuring the relationship that exists between a variable and lagged values of itself. 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 (Tim, 2023).

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).

According to existing literature, there are three major ways of statistically controlling an autocorrelated process. The first is the model-based approach where a time series model is fitted to the process data thereby obtaining residuals that are *i.i.n.d*, afterwards, the residuals are used as process control data which can be plotted using the traditional control charts (Siaw *et al.*, 2015). The second method is the one in which the control limits of the traditional control charts are adjusted to account for autocorrelation, and the third which is less popular is the non-parametric method which does not require any model fitting strategies.

Also, the potential law which governs most autocorrelated data is the Geometric Brownian Motion as explained in Siaw et *al.* (2013), and these GBM is an Autoregressive of order 1 process. With this point as a motivation, Logarithm return model (LRM) was introduced by Siaw et *al.* (2013), to overcome the problem of sophistication which usually comes with the use of ARIMA models for removing autocorrelation in a time series.

In Statistical Process Control, the goal is to produce goods and services which are as identical as possible, so variation is expected to be minimal, as explained by Muhammad *et al.* (2023). According to literature, logarithmic returns are approximately equal to the arithmetic ones when the rate of return is not much, and so this factor is the motivation for this work.

Arithmetic return model (ARM) was introduced by Chajire *et al.* (2024) as a transformation of an Autoregressive process of order 1 (AR(1)) for removing autocorrelation from a process which exhibits geometric brownian motion (GBM) and compared to LRM of Siaw *et al.* (2022). The major goal in this work is to find out the strength of ARM for control and monitoring of autocorrelated data which exhibits the geometric Brownian motion (GBM).

MATERIALS AND METHODS Arithmetic Return Model (ARM)

The model used was achieved by converting the Autoregressive process of order 1 (AR (1)) to its Arithmetic Return (AR) and the resulting model was used for removal of autocorrelation, and control of some processes.

To consider the Arithmetic Return Model (ARM) for time dependent data, transformation of Arithmetic Return at time t (AR_t) is needed and shown as

$$AR_t = \frac{x_t - x_{t-1}}{x_{t-1}} \tag{1}$$

Consider an AR(1) process which can be written as; $AR_t = C + \Theta AR_{t-1} + e_t$ (2)

where *C* is the intercept, Θ the slope in AR (1) model and $e_t \sim N(0, \sigma^2)$

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

$$\frac{\tilde{X}_{t} - X_{t-1}}{X_{t-1}} = C + \Theta\left(\frac{X_{t-1} - X_{t-2}}{X_{t-2}}\right) + e_t$$
(3)
thus,
 $X_{t} = 1 - C + \Theta\left(\frac{X_{t-1} - X_{t-2}}{X_{t-2}}\right) + e_t$

$$\frac{x_{t}}{x_{t-1}} - 1 = C + \Theta\left(\frac{x_{t-1}}{x_{t-2}} - 1\right) + e_{t}$$

$$\frac{x_{t}}{x_{t-1}} = C + 1 + \Theta\left(\frac{x_{t-1}}{x_{t-2}}\right) - \Theta + e_{t}$$

$$\frac{x_{t}}{x_{t-1}} = [C + 1 - \Theta] + \Theta\left(\frac{x_{t-1}}{x_{t-2}}\right) + e_{t}$$

$$\frac{x_{t}}{x_{t-1}} = \mathcal{B} + \Theta\left(\frac{x_{t-1}}{x_{t-2}}\right) + e_{t}$$
(4)

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

Therefore, Equation (4) is the Arithmetic Return model (ARM). The residuals were obtained using Equation (5) $\hat{e}_t = X_t - \hat{X}_t$ (5)

The logarithmic return model (LRM) used for comparison is the one proposed by Siaw *et al.* (2022) as shown in Equation (6)

$$X_{t} = e^{C} \cdot X_{t-1} \cdot \left(\frac{X_{t-1}}{X_{t-2}}\right)^{\theta} + \mathcal{E}_{t}$$
(6)

The Durbin-Watson (d) Test for Autocorrelation

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

This test was carried out on both sets of data before the main analysis, to confirm for the presence of positive autocorrelation. The obtained residuals from both the (ARM) and (LRM) models were also tested to check if autocorrelation is removed from both datasets after the transformations.

The null and alternative hypotheses used for the Durbin-Watson test are:

 H_0 : Positive autocorrelation does not exist among the process residuals

*H*₁: Positive autocorrelation exists among the process residuals

Test statistic:

 $d = \frac{\sum_{t=2}^{T} (e_t - e_{t-1})^2}{\sum_{t=1}^{T} e_t^2}$ (7) where *T*: The total number of observations

 e_t : The tth residual from the regression model (Zach, 2021).

Decision rule for Durbin-Watson test for the presence of positive autocorrelation

At significance level α (alpha), the DW test statistic was compared at lower and upper critical values as presented by CFI (2024).

If $d > d_U$, there is no statistical evidence that positive autocorrelation exists in the data

If $d < d_L$, there is statistical evidence that positive autocorrelation exists in the data.

If $d_L < d < d_U$, the test is inconclusive.

The EWMA Residual Control Chart

For the EWMA residual control chart, the EWMA statistic as defined by (Aytaç, 2020), is

$$\Omega_t = \lambda e_t + (1 - \lambda)\Omega_{t-1} \tag{8}$$

Where λ is a constant (smoothening parameter) $0 \le \lambda \le 1$ which determines the depth of memory of the EWMA. $\Omega_{t's}$ are the EWMA's at time *t*

 e_t are the errors at time t

In order to detect small shifts, small values for λ are preferred and vice versa. The initial value Ω_0 can be taken equal to the in-control mean of the residuals, which is zero (0). Just as in the case of the traditional EWMA chart, the control limits of the EWMA residual chart are initially not constant, they reach a constant value after some time (Aytaç, 2020). These steadystate control limits will be calculated as follows:

UCL =
$$\mu + 3\sigma_{e}\sqrt{\frac{\lambda}{2-\lambda}}$$
 (9)
CL = μ (10)

$$LCL = \mu - 3\sigma_{e} \sqrt{\frac{\lambda}{2-\lambda}}$$
(11)

Where.

 σ_e = standard deviation computed from the errors μ = in - control mean of residuals λ = the smothening parameter

RESULTS AND DISCUSSION

Durbin-Watson (*d*) Test for the Presence of Positive Autocorrelation

Durbin-Watson test was carried out on the simulated process and furnace temperature data obtained from Salleh *et al.* (2018), to confirm the presence of positive autocorrelation before the proposed model transformation was done on each of them. This section provides initial Durbin-Watson test results of the three datasets using Equation (7) and their interpretations.

From the simulated process data,

$$d = \frac{12180.95534}{10200.06222} = 1.17126$$

From the Durbin-Watson table at 5% significance level, n = 200, k = 1 (no of predictor variables) the lower and upper values are $d_L = 1.758$ and $d_U = 1.779$. DW test result obtained from the residuals of the simulated process data is 1.17126. Here $d < d_L$ indicating that positive autocorrelation exists in the simulated data. Therefore, the null hypothesis which precedes Equation (7) was rejected here and conclusion was made that positive autocorrelation exists among the residuals of the data which needs to be removed before statistical monitoring of the process.

From the furnace temperature data,

$$d = \frac{36.60311}{24.32967} = 1.504464$$

From the Durbin-Watson table at 5% significance level, n = 80, k = 1 (no of predictor variables) the lower and upper values are $d_L = 1.611$ and $d_U = 1.662$. DW test value obtained from the residuals of the furnace temperature data is 1.5045. Here $d < d_L$ indicating that positive autocorrelation exists in the furnace temperature data. Therefore, the null hypothesis which precedes Equation (7) was also rejected, and conclusion was made that positive autocorrelation exists among the residuals of the furnace temperature data which needs to be removed before statistical monitoring of the process.

Results of Analysis from the ARM and LRM of Simulated Process and Furnace Temperature Datasets The two datasets used were transformed to ARM and LRM and the results from the analysis are presented in this section. Table 1 and 2 presents results for both processes without any transformation carried out on them, alongside the ARM and LRM transformation results.

Model	Slope	Intercept	\mathbb{R}^2	Correlation	d
Simulated process	0.516556	86.26962	0.215098	0.517134	1.171261
ARM	-0.44707	4.21E-04	0.203957	-0.45162	2.86818
LRM	-0.44607	6.57E-05	0.173363	-0.45079	2.143208

The fitted time series equation from the untransformed simulated process data is $\hat{Y} = 86.2606 \pm 0.5166 Y$ (12)

$Y_t = 86.2696 + 0.5166X_t$	(12)
The ARM fitted to the simulated process dataset i	is,

$$\begin{pmatrix} \frac{X_t}{X_{t-1}} \end{pmatrix} = 4.21 \times 10^{-4} + (-0.4471 \left(\frac{X_{t-1}}{X_{t-2}} \right))$$
(13)
The LRM fitted to the simulated process data is,
$$X_t = e^{6.57 \times 10^{-5}} \cdot X_{t-1} \cdot \left(\frac{X_{t-1}}{X_{t-2}} \right)^{-0.4461}$$
(14)

MODEL	Slope	Intercept	R ²	Correlation	d
Furnace Temperature	0.72388	435.98496	0.50021	0.70725	1.50446
ARM	0.14546	-4.83×10^{-6}	0.02173	0.14742	1.7848
LRM	0.14545	-4.89×10^{-6}	0.02173	0.14741	2.00465

The fitted time series equation from the untransformed furnace temperature data is

$$\begin{split} \hat{Y}_t &= 435.9850 + 0.7239 X_t \quad (15) \\ \text{The ARM fitted to the furnace temperature data is,} \\ & \left(\frac{x_t}{x_{t-1}}\right) = -4.8315 \text{ x } 10^{-6} + 0.145460909 \left(\frac{x_{t-1}}{x_{t-2}}\right) (16) \end{split}$$

The LRM fitted to the furnace temperature data is,

$$X_{t} = e^{-4.8926 \times 10^{-6}} \cdot X_{t-1} \cdot \left(\frac{X_{t-1}}{X_{t-2}}\right)^{0.14545076}$$
(17)

Time Series Plots of the Simulated Process Data

The time series plots of the simulated process data, ARM and LRM transformed data, all against time (in hours) are presented in Figure 1.



Figure 1: Time series plot of the simulated process data, ARM and LRM transformations.

The random fluctuations and sudden drifts of this data over time shows that the simulated process is GBM. The markovian nature of the plot is also indicative that the process is GBM. Figure 1 shows how the ARM and LRM performed similarly in giving approximately the same fitting to simulated process data. This indicates here, that the ARM competes well with the existing LRM is terms of fitting of the simulated process dataset.



Time Series Plots of the Furnace Temperature Data

Figure 2: Time series plots of the furnace temperature, ARM and LRM transformations.

The change in the values of the furnace temperature from one time period to the next is a random variable, and the tendency of the data to move randomly over time but with systematic upward and downward trend is indicative of GBM. The process also proves to be GBM for its time continuous and Markovian nature. The performance of both the ARM and LRM in fitting of the furnace temperature dataset is well distinguished here. The ARM forecast to the furnace temperature gives a better fit to the furnace temperature dataset, with the trend line from the ARM closely beside the untransformed furnace temperature data points.

Exponentially Weighted Moving Average (EWMA) Control Charts

EWMA control charts were constructed to monitor the simulated data and furnace temperature data obtained from

Salleh *et al.* (2018) around a specified target, using the Arithmetic Return Model (ARM) after each dataset was studied and confirmed to exhibit the GBM nature. Varying values of the smoothening parameter λ (0.25, 0.5 and 0.7) were used for the control and observed.

EWMA Control Charts for the Simulated Process Data Residuals

EWMA residual control charts were constructed using the ARM and the LRM to monitor the simulated process data residuals around a specific target value (μ) and to visualize how residuals from both models relate with the EWMA at varying values of the smoothening parameter (λ).



Figure 3: EWMA control chart for ARM of simulated process data, $\lambda = 0.25$

At $\lambda = 0.25$, UCL= 6.1609, CL= 0.0368, LCL= -6.2345, Figure 3, shows that some of the simulated process residuals fall outside of the control limits, and this can be as a result of the strictness of the smoothening parameter at $\lambda = 0.25$, then higher λ value needs to be selected to keep the process in state

of statistical control. Control limit values for the ARM are similar to those of the existing LRM and so, similar performance was recorded for this dataset.

FJS



Figure 4: EWMA control chart for LRM of simulated process data, $\lambda = 0.25$ At $\lambda = 0.25$, UCL= 6.2160, CL= 0.00022, LCL= -6.2205, the LRM performs similarly to the EWMA of the ARM with controls limits quite similar.



Figure 5: EWMA control chart for ARM of simulated process data, $\lambda = 0.5$

At $\lambda = 0.5$, UCL= 10.6979, CL= 0.0368, LCL= -10.7716, all of the simulated process data residuals fall within the control limits with an increase in the value of the smoothening parameter. The increase in control limits is indicative of an increase in the λ value. This increases the probability that

more of the process residuals fall within the acceptance region of the control chart. The control chart for this dataset works well with values of λ ranging from 0.5 to 1.



Figure 6: EWMA control chart for LRM of simulated process data, $\lambda = 0.5$

At $\lambda = 0.5$, UCL= 10.7681, CL= 0.0022, LCL= -10.7726, the LRM performed similarly to the ARM, having similar values of the control limits and most of the data points falling within the upper and lower control limits.





Figure 7: EWMA control chart for ARM of simulated process data, $\lambda = 0.7$ At $\lambda = 0.7$, UCL= 16.3607, CL= 0.0368, LCL= -16.4344, all data points fall within the statistical control limits indicating that the process is in the state of statistical control, haven taken care of the autocorrelation problem.



Figure 8: EWMA control chart for LRM of simulated process data, $\lambda = 0.7$ At $\lambda = 0.7$, UCL= 16.4497, CL= 0.0022, LCL= -16.4542, performance of the control chart from the existing LRM is also similar to that of the ARM.

EWMA Control Charts for Furnace Temperature Data Residuals

EWMA residual control charts were constructed using the ARM and the LRM to monitor the furnace temperature data

residuals around a specific target value (μ) and to visualize how both models respond to the control charts for the residuals at varying values of the smoothening parameter (λ).



Figure 9: EWMA control chart for ARM of furnace temperature, $\lambda = 0.25$

At $\lambda = 0.25$, UCL= 0.0713, CL= 0.0310, LCL= -0.0092, four of the data points clearly fall outside of the upper control limit and the cause for such drift is credited largely to random sources and not as a result of autocorrelation in the process. Also, in comparison to Figure 16, the control chart of the ARM performed better than that for the LRM in faster

detection of out-of-control signals, with four points clearly detected outside and only three points detected outside the upper control limits of the ARM and LRM control charts respectively.



Figure 10: EWMA control chart for LRM of furnace temperature, $\lambda = 0.25$ Here, $\lambda = 0.25$, UCL = 0.0716, CL= 0.0321, LCL = -0.0075, and only three of the data points clearly fall outside the upper control limit. This clearly shows that the ARM for this set of data detects out of control signal faster than the LRM.



Figure 11: EWMA control chart for ARM of furnace temperature, $\lambda = 0.5$

At $\lambda = 0.5$, UCL = 0.1007, CL= 0.0310, LCL= -0.0386, two of the data points fall outside of the upper control limit also. Number of points outside the upper control limit have decreased with an increase in the value of the smoothening

parameter (λ). Also, the ARM performed better than the LRM here in faster detection of the second point falling outside the upper control limit in comparison with Figure 18.



Figure 12: EWMA control chart for LRM of furnace temperature, $\lambda = 0.5$ At $\lambda = 0.5$, UCL= 0.1006, CL = 0.0321, LCL = -0.0364 also, number of points falling outside the upper control limit also have decreased to two with an increase in the λ value and the ARM still detected the out of control point faster than the LRM.



Figure 13: EWMA control chart for ARM of furnace temperature, $\lambda = 0.7$

At $\lambda = 0.7$, UCL= 0.1375, CL= 0.0310, LCL= -0.0754, all points fall within the control limits indicating that all the process residuals are within limits of statistical control. The control charts for this dataset works well with λ values ranging from 0.6 to 1 for the ARM.



Figure 14: EWMA control chart for LRM of furnace temperature, $\lambda = 0.7$ At $\lambda = 0.7$, UCL = 0.1368, CL = 0.0321, LCL = -0.0726, all points fall within control limits indicating that this process is in a state of control.

Arithmetic Return Model (ARM) was proposed in this work for Statistical Process Control (SPC) of autocorrelated processes exhibiting Geometric Brownian Motion (GBM). The ARM successfully removed autocorrelation from all the datasets analyzed, demonstrating greater ease of understanding and simplicity in implementation compared to the existing Logarithmic Return Model (LRM). Additionally, the ARM exhibited a faster detection capability for out-ofcontrol signals in SPC of autocorrelated data, making it a more appealing choice than the LRM. The ARM effectively fits SPC datasets for monitoring and control, significantly reducing process variation. Future research could extend the application of the ARM to real life-life processes for monitoring and controlling autocorrelated data exhibiting GBM characteristics, further elucidating the benefits of the ARM. Another potential extension is exploring the applicability of the model to multivariate process control of autocorrelated data exhibiting GBM behavior. in this work which exhibit the GBM nature, with much ease in understanding and simplicity in implementation than the existing LRM.

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/

Chajire, B. P., Dike, I. J., Abdulkadir S. S., and Torsen, E. (2024). Arithmetic Return model for removing autocorrelation from Statistical process control data exhibiting Geometric Brownian Motion. *BIMA Journal of Science and Technology*. 8(1B), 281-286. https://doi.org/10.56892/bima.v8i1.634

Chao-Wen, L., and Marion, R.R. (2001). Cusum Charts for Monitoring an Autocorrelated Process. *Journal of Quality Technology*. 33(3), 316-334. <u>htt</u> ps://doi.org/10.1080/00224065.2001.11980082

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

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

Muhammad N. A., Garba, N. N., and Michael, A. O. (2023). Quality Control Analysis Of Diagnostic Radiology Equipment In 44 Nigerian Army Reference Hospital Kaduna, Kaduna State, Nigeria. *FUDMA JOURNAL OF SCIENCES*, 2(4), 61 - 66. Retrieved from https://fjs.fudutsinma.edu.ng/index.php/fjs/article/view/1319

Rupali, K. and Vikas, G. (2022). Estimation of the Change Point in the Mean Control Chart for Autocorrelated Processes. *RT&A*. 17(3), 69.

Salleh, R. M., Zawawi, N. I., Gan Z. F., and Nor, M. E. (2018). Autocorrelated process control: Geometric Brownian Motion approach versus Box-Jenkins approach. *Conference Series: Journal of Physics*. 995(012039), 1-11. https://doi.org/10.1088/1742-6596/995/1/012039

Siaw, L. L., and Djauhari, M. A. (2013). Monitoring Autocorrelated Processes: A Geometric Brownian motion process approach. <u>https://www.researchgat</u> <u>e.net/publication/260829536</u> <u>https://doi.org/10.1063/1.4823976</u>

Siaw, L. L., Djauhari, M. A., and Mohamad, I. (2015). Modeling Autocorrelated Process Control with Industrial Application. <u>https://www.researchgat</u> <u>e.net/publication/282928394</u> <u>https://doi.org/10.1109/IEEM.2014.7058614</u>

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. *Investopedia*. Retrieved on 14th August 2023 from https://www.investopedia.com/terms/a/autocorrelation.asp

Zach (2021) The Durbin-Watson Test: Definition and example. *STATOLOGY*. Retrieved on 9th May, 2023 from <u>https://www.statology.org/durbin-watson-test</u>



©2025 This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 International license viewed via <u>https://creativecommons.org/licenses/by/4.0/</u> which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is cited appropriately.