International Journal of Mathematics and Statistics Invention (IJMSI)
E-ISSN: 2321 – 4767 P-ISSN: 2321 - 4759
www.ijmsi.org Volume 3 Issue 6 || September. 2015 || PP-15-23

Assessing the Effects of Autocorrelation on the Performance of Statistical Process Control Charts

Aminu Ahmad Magaji[1*], Abubakar Yahaya[2], Osebekwin Ebenezer Asiribo[3]
[1,2]Department of Mathematics, Ahmadu Bello University, Zaria - Nigeria
[3]Department of Statistics, Federal University of Agriculture, Abeokuta – Nigeria

**Corresponding author email: aamagaji1@gmail.com*

Abstract: *This study discusses the applications of univariate control chart especially in assessing its performance on autocorrelated data comprising 100 indicators obtained based on a chemical process viscosity measured at two minutes intervals. The graph of the autocorrelation function based on the original data showed the presence of autocorrelation thereby necessitating data remodeling in order to attain independency as well as normality. The data was modeled in order to determine the effectiveness of parameter estimation and Box-Jenkins methodology was used to determine the estimates of the parameters of the identified model and each parameter was statistically tested for significance. The time series plot, autocorrelation and partial autocorrelation functions suggested some models for selection, but the Akaike Information Criterion was used to select the model that gives the best fit for the data. From the family of identified models, ARIMA (2.0.0) model was found to be most adequate due to its ability to capture the presence of autocorrelation in the data. The adequacy of the chosen model was subsequently checked using both the residual analysis and Ljung-Box statistics.*

Keywords - *ARIMA model, Autocorrelation, EWMA Control Charts, Statistical Process Control.*

I. Introduction

Since before the industrial revolution, the process of monitoring and controlling quality of products in industries has been the rallying point of most experienced artisans and craftsmen of the time, and this is made possible through interactions with customers aimed at assessing their (customers') satisfactions on a particular product or services. According to Duncan [5], statistical quality control (SQC) is undoubtedly as old as the history of the industry itself. Steiner [17] posits that SQC was first introduced in the 1920s by the Inspection Department at Telephone Laboratories under the combine efforts of Walter A. Shewhart, Harold F. Dodge, Donald A. Quarles and George D. Edwards purposely for non-continuous manufacturing process that generate independent data and since then lots of charts have been developed for different process data in order to improve the quality of the products or services. SQC comprises some techniques of analyzing the processes involving comparison of performances, verifying and studying deviations as well as analyzing the processes again and again with the sole aim of achieving the best performance of machinery's and/or operators (Montgomery [10]).

The fundamental assumption for statistical process is that, the data generated by the in-control process are identically and independently distributed normal variables with mean, μ and standard deviation,σ. However, in continuous process the assumption of independency is not always assured, particularly when successive units are related to previous one; hence, standard control charts may exhibit an increased frequency of false alarms. The data will then tend to be autocorrelated as the process time is longer than the time between samples collections (Nembhard and Nembhard [12]).

Control chart as a major and primary tool for statistical process control (SPC), can only detect assignable causes, which need to be eliminated through some managerial, operator, and engineering actions. Control charts consist of three parallel lines: Central Line (CL), Upper Control Limit (UCL), and Lower Control Limit (LCL). Control chart can be subdivided into variable and attribute control charts and both are sensitive to the presence of autocorrelation in the process data even at low level. The degree of dependency between set of observations normally interrupt the chart's properties which may eventually leads to signaling some false alarm and shorter average run length. Autocorrelation between observations results from different factors right from the operator or the process itself. According to Russo *et al*. [13], Shewhart control chart is not much sensitive in detecting small and moderate shifts in the process parameters and to overcome this insensitivity western electric run rules can be applied. However, as the sensitivity to detect assignable causes of variation is increased using run rules, the risk of sending false alarms from the chart properties also increases. Alternatively, one can apply

Exponential Weighted Moving Average (EWMA) or Cumulative Sum Control Chart (CUSUM) charts. The main advantages of employing these lies in their ability to quickly detect relatively small shift in the process mean, and this detection is significantly quicker than Shewharts control charts.

Russo *et al*. [13] observed that, the quality of a product manufactured by a given process is normally subjected to two types of variations. First, no matter how good the design of a process is; there are some special or natural causes that may affect the process in an unpredictable manner and any process that operates under the influence of relatively special causes is considered to be in control and acceptable. Secondly, there are common or assignable causes whose variation affects all the individual values of the process and any process that performs in the presence of assignable causes is considered to be an out-of-control process, particularly if the resultant variations are significant compared to the specification limits of the product.

Karaoglan and Bayhan [7] argued, when there is significant autocorrelation in the process data, traditional control chart cannot be applied directly without some modifications which can be achieved through three general approaches, namely: (I) modelling the process observations using Autoregressive Integrated Moving Average (ARIMA) models and then applying traditional control charts to process the residuals; (II) adjusting the standard control limits in the traditional control charts in order to account for the autocorrelations from the process observation; (III) eliminating autocorrelation from the process observations using Engineering controllers. Kandananond [6] noted that, an effort to adjust a stable process in order to compensate for an undesirable disturbances may temper with the process and may also lead to more variations, and so the best practice is to integrate forecasting models into the traditional SPC tools.

Several quality control researchers [3, 4, 5, 6, 8, 14, 18] studied the behavior of control chart performance in the presence of autocorrelation, and the fundamental observation was that autocorrelation affects control chart performance of both variable as well as attribute control charts results. According to Alwan and Roberts [3], more than 85% of industrial process control applications display incorrect control limit due to the autocorrelation of the process observations thereby violating the fundamental assumption of Shewhart control chart. According to Sanders and Dan Reid [14], variables control chart can be used to monitor quality characteristic that can be measured and having a continuous scale. Descriptive statistics such as measures of central tendency and variation can as well be very helpful in describing certain characteristics of a product and a process. A control chart for attribute on the other hand, can be used to monitor characteristics that are discrete in nature and can be counted and sometimes this can be achieved with a simple 'yes' or 'no' decision. Nembhard and Nembhard [12] considered the effect of autocorrelation in attribute control chart and the study is related to injection modeling production target produced by various models of leak proof plastic containers.

Shey and Shin [15] posits that, it is more advisable to use EWMA control charts in detecting small shifts than Shewhart control charts especially for observations drawn from AR (1) model with random error. However, it was argued that, Generalized Weighted Moving Average (GWMA) control charts are more powerful than their EWMA counterparts in detecting small shift in the process mean and variable. This study, motivated by the observance of fluctuation in some production processes, is aimed at providing better understanding on how to minimize the production and consumption of substandard products. The study is believed to shed more light on the applicability of the ARIMA modeling techniques in SPC using chemical process viscosity data obtainable in Montgomery and Johnson [9]; and this can be attained by exploring the effect of autocorrelation on control chart performance and proposing an approach for constructing control charts that removes the impact of autocorrelation on observations. The models developed in this study are expected to have some degree of accuracy which may be useful in process setting, monitoring and decision making.

II. Materials and methods

The data used in this study was obtained from a research conducted by Montgomery and Johnson [9] comprising 100 indicators of chemical process viscosity measured and recorded at two-minute intervals. The essence is to assess control chart performance and obtain a prediction model that measures and as well removes the impact of autocorrelation on the data. EWMA control charts were applied to the data as well as the residuals; and the two results were compared to determine effect of autocorrelation on control chart performance. The data analysis was conducted using a statistical software package - MINITAB.

2.1. EWMA Control Charts

Roberts introduced EWMA control charts in 1959 as alternative to the conventional Shewhart control charts especially when one is interested in detecting a small shift in the process. It was established that, CUSUM and EWMA charts are suitable and good candidates for detecting small process shift; for further details, consult Shu and Tsung [16]. EWMA control charts are approximately equivalent to their CUSUM counterparts and in some cases are easier to set up and operate. EWMA control charts are usually applied in industrial process set up where the products comes one by one. The EWMA statistic is given by:

$$Z_t = \lambda x_t + (1 - \lambda) Z_{t-1} \tag{1}$$

Where Z_t is the moving average at time *t*, λ is the smoothing constant taking values in the interval (0, 1) and mostly chosen between 0.05 and 0.3. Karaoglan and Bayhan [7] as well as Lucass and Saccucci [8] developed tables that serve as a guide on how to select an optimal value for λ.

According to Ajit and Dutta [1], EWMA control limit for independent and normally distributed data are given by:

$$UCL = \mu_0 + L\sigma\sqrt{\frac{\lambda}{2-\lambda}\left[1-(1-\lambda)\right]^{2t}} \tag{2}$$

$$CL = \mu_0 \tag{3}$$

$$LCL = \mu_0 - L\sigma\sqrt{\frac{\lambda}{2-\lambda}\left[1-(1-\lambda)\right]^{2t}} \tag{4}$$

Where *L* is the number of standard deviations from the CL, if the observations x_t are independent random variables with variance σ^2, then the variance of Z_t will be given as:

$$\sigma^2_{2t} = \sigma^2\left(\frac{\lambda}{2-\lambda}\right)\left[1-(1-\lambda)^{2t}\right] \tag{5}$$

The term $\left[1-(1-\lambda)^{2t}\right]$ approaches unity as *t* get larger, so after several time periods, the control limits will eventually approach:

$$UCL = \mu_0 + L\sigma\sqrt{\frac{\lambda}{2-\lambda}} \tag{6}$$

$$LCL = \mu_0 - L\sigma\sqrt{\frac{\lambda}{2-\lambda}} \tag{7}$$

Douglas [7] observed that in many situations, the sample size used for process control is *n = 1*, meaning that the sample consist of an individual unit only.

2.2 Time series models

One of the methods of removing autocorrelation from a given process data is the concept of modeling using ARIMA model. ARIMA control charts are normally used to find an appropriate time series model, and the residual from the model are applied to control charts. An ARIMA of order *p*, *q*, *d* {i.e. ARIMA (*p*, *d*, *q*)} is chosen based on the characteristics shape of estimated autocorrelation function (EACF) and estimated partial autocorrelation function (EPACF). For instance, Russo *et al*. [13] suggested Box Jenkins and Reinse general shape of the model {i.e. ARIMA (*p*, *d*, *q*)} is given by:

$$\Phi_p(B)\nabla^d X_t = \Theta_q(B)\varepsilon_t \tag{8}$$

in which

$\Phi_p(B) = (1 - \phi_1 B - \phi_2 B^2 ..... - \phi_p B^p)$; $\Theta_q(B) = (1 - \theta_1 B - \theta_2 B ..... - \theta_q B^q)$

$\mathrm{B} = \frac{X_{t-1}}{X_t}$; $\nabla = \frac{X_t - X_{t-1}}{X_t} = 1 - \mathrm{B}$

Where B is backward shift operator, ∇ is a backward difference operator, $\Phi_p(\mathrm{B})$ is autoregressive polynomial of order *p* and $\Theta_q(\mathrm{B})$ is moving average polynomial of order *q*. The observation at time period *t*, is denoted by X_t. Lastly ε_t denote independent white noise at time period *t* having normal distribution with mean 0 and variance σ^2.

The most commonly ARIMA model used in applications can be summarized in the following equation:

$$X_t = \xi + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + \varepsilon_t \qquad (9)$$

Where X_t is a p^{th} order autoregressive process or *AR (P)* process for short. $\phi_1, \phi_2, \cdots, \phi_p$ denote estimated coefficients and each lies within the interval of $-1 < \phi < 1$, and ξ is an unknown constant, ε_t is an error term at time *t.*

For instance, a first order autoregressive {*AR (1)*} model is given by:

$$X_t = \xi + \phi_1 X_{t-1} + \varepsilon_t \qquad (10)$$

In this process, the estimated/fitted values $\hat{X}_t$ are subtracted from the sample data X_t, the residuals that are approximately normal and independently distributed with mean 0 and constant variance $\sigma^2$are of the form:

$$\varepsilon_t = X_t - \hat{X}_t$$

and at this stage the residuals can be applied to the conventional control chart. The estimators of process mean and variance for AR models are given as:

$$\mathrm{E}(X) = \mu = \frac{\xi}{1 - \sum_{j=1}^{p} \phi_j} \qquad (11)$$

$$Var(X) = \gamma_0 = \frac{\sigma_\varepsilon^2}{1 - \sum_{j=1}^{p} p_j \phi_j} \qquad (12)$$

Thus, a first order moving average model is given as:

$$X_t = X_{t-1} + \varepsilon_t - \theta_1 \varepsilon_{t-1} \qquad (13)$$

where X_t is the observation at time *t*, θ is the moving average parameter whose range of values lie within the interval (-1 < θ < 1), , ε_t is a random error term at time *t.* The dependency between X_t and X_{t-1} is captured in $P_1 = \frac{\theta}{(1+\theta^2)}$ and is zero at all other lags, the estimators of process mean and variance for MA models are thus computed as:

$$\mathrm{E}(X) = \mu \qquad (14)$$

$$Var(X) = \gamma_0 = \sigma_\varepsilon^2 \sum_{j=0}^{q} \theta_j^2, \theta_0 = 1 \qquad (15)$$

The Autoregressive Moving Average (ARMA) model which combines the autoregressive and moving average parameters, the first order mixed model as:

$$X_t = \xi + \phi X_{t-1} + \varepsilon_t - \theta\varepsilon_{t-1} \tag{16}$$

The estimators of process mean and variance for ARMA model are:

$$\mathrm{E}(X) = \mu = \frac{\xi}{1 - \sum_{j=1}^{p} \phi_j} \tag{17}$$

$$Var(X_t) = \gamma_0 = \sum_{j=1}^{p} \phi_j \gamma_j - \theta_1 \gamma_{X\varepsilon}(-1) - \cdots - \theta_q \gamma_{X\varepsilon}(-q) + \sigma_\varepsilon^2 \tag{18}$$

Where σ_ε^2 is the variance of the residuals, ϕ_j the order parameter *j* of AR or ARMA model, p_j is the correlation coefficient of lag *j* and γ_j auto-covariance of lag *j*.

ARIMA combines both Autoregressive and Moving Average parameters and also include differencing, the model is given by:

$$\Delta_d X_t = \mu + \Phi_1 \Delta_d X_{t-1} + \Phi_2 \Delta_d X_{t-2} + \cdots + \Phi_p \Delta_d X_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_p \varepsilon_{t-p} \tag{19}$$

The above model comprises three types of parameters including autoregressive parameter (*p*), number of differencing passes (*d*), and moving average parameter (*q*), the model can be rewrite as ARIMA (*p*, *d*, *q*). Modeling using Box and Jenkins methodology establishes certain steps before the best model can be determined. These include identification, estimation, checking and forecast. ARIMA input series need to be a stationary process.

2.2.1. Autocorrelation

Autocorrelation measures the dependency between series of observations collected at difference time intervals. The graphical representation of autocorrelation function is called Correlogram. Let X_t denote a time series, the ratio between the covariance (X_t, X_{t-k}) and variance (X_t) defines as simple autocorrelation coefficient (r_k), while he sequence of r_k value is called autocorrelation function simple (AFS). The autocorrelation coefficient simple between X_t and their X_{t-k} lagged values are defined by Russo *et al*. [13] as follows:

$$r_k = \hat{\rho}_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0} = \frac{\sum_{t=1}^{N-k} (X_t - \bar{X})(X_{t+k} - \bar{X})}{\sum_{t=1}^{N} (X_t - \bar{X})^2} \tag{20}$$

Where $Cov(X_{t-1}, X_t, \cdots, X_{t+k})$ is the covariance of observations that are *k* periods apart, X_k is the auto-covariance of lag *k*, X_0 is the auto-variance of lag *k* = 0, and *N* is the total number of observations in the dataset.

The standard error, s_{ek} at lag *k* is given by:

$$s_{e_k} = \begin{cases} \sqrt{\dfrac{1}{n}} & for \ k = 1 \\ \sqrt{\dfrac{1}{n}(1 + 2\sum_{i=1}^{k-1} r_i^{2})} & for \ k > 1 \end{cases} \qquad (21)$$

III. Result and discussions

In the previous sections, we described the effect of autocorrelation on control chart performance and modeling techniques using Box-Jenkins methodology. As for this section, we present an analysis of a numerical data on chemical process viscosity measurements. After comparing the ACF with the PACF; we discovered that, ARIMA (2, 0, 0) provided the best fit having the lowest value of Akaike Information Criterion (AIC). The residuals obtained from the model were also analyzed to confirm the removal of autocorrelation from the data and the analysis was performed using MINITAB version 15 package.

3.1. Time series plot of data

The following figure (Fig. 1) shows time series plot of the original data and some runs which indicates the dependency embedded in the dataset between the set of observations due to the presence of autocorrelation.

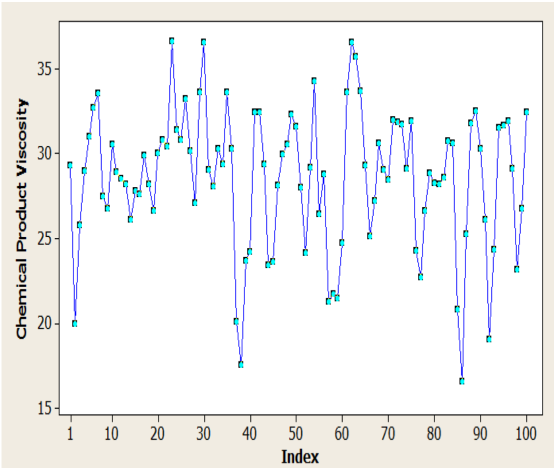

Fig. 1: Time series plot of the original data before modifications.

3.2. Results of Autocorrelation Function

Figure (2) and (3) illustrate the graphs of ACF and PACF and both indicate the presence of autocorrelation in the data. It is clear that, one can assess lags significantly different from zero and the series is not a white noise indicating the need for data modeling.

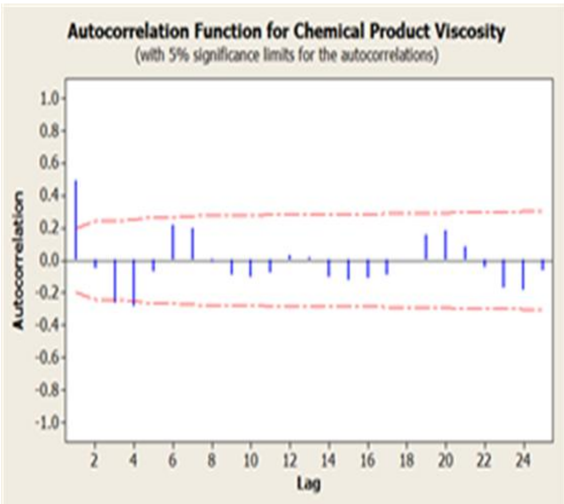

Fig. 2: Autocorrelation function (ACF) of the data.

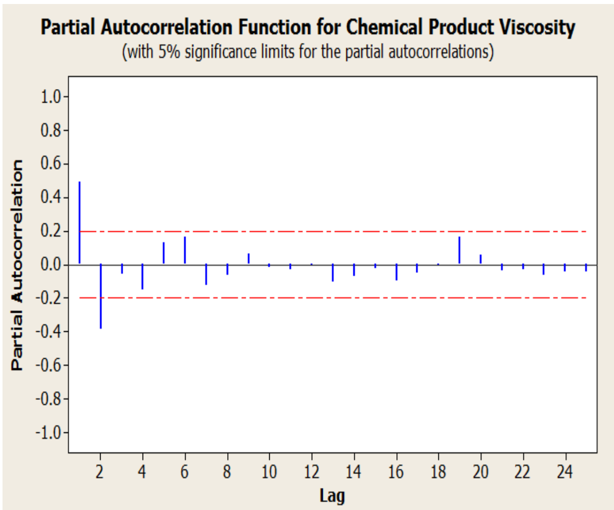

Fig. 3: Partial autocorrelation function (PACF) of the data

3.3. Result of removing autocorrelation

To remove autocorrelation effect from a given data set; Montgomery and Johnson [9] recommended data modeling as well as control charting the residuals directly. Table 1 below gives information about the final estimates of the parameters showing how the ARIMA (2, 0, 0) model fits in well.

Table 1: Final estimates of parameters

Type	Coef	SE Coef	T	P
AR 1	0.7182	0.0923	7.78	0.000
AR 2	-0.4338	0.0922	-4.70	0.000
Constant	20.5022	0.3281	62.49	0.000
Mean	28.6512	0.4585		

Number of observations = 100
Residuals: SS = 1043.73, MS = 10.76, DF = 97 (back forecast excluded).
It can easily be observed from Table 1 that, the values of the estimates of the parameters ϕ_1 and ϕ_2 are as given below:

ξ = 20.5022, ϕ_1 = 0.7182, ϕ_2 = -0.4338

Thus, the resultant AR (2) model is as follows: $\hat{X}_t = 20.5022 + 0.7182 X_{t-1} - 0.4338 X_{t-2}$
And now the residuals are independently and identically distributed normal variables.

3.4. Results of normal probability as well as residuals time series plots.

Figures 4 and 5 respectively show a normal probability plot of the residuals as well as residual time series plots in which both showcase the fitness of the ARIMA (2, 0, 0) in fitting the data well.

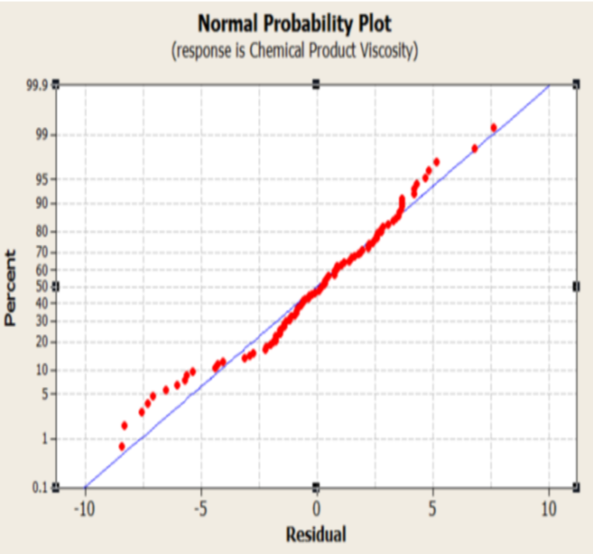

Fig. 4: Normal probability plot of the residual

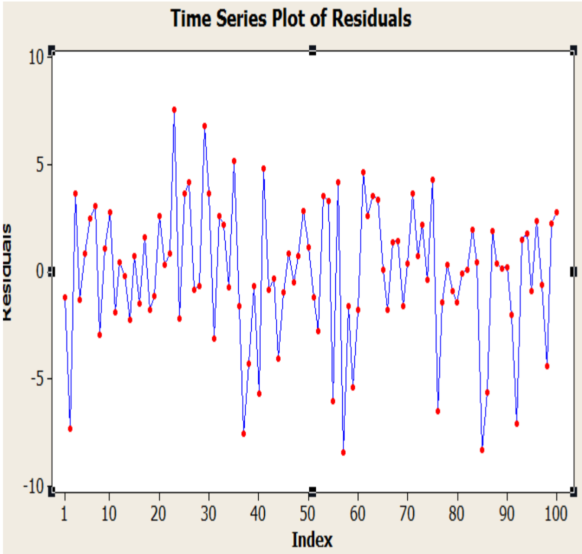

Fig. 5: Time series plot of the residual.

3.5. Result of ACF and PACF

After fitting the model, there is the need to further check ACF and PACF of the residuals obtained. The model is considered adequate and fit if the autocorrelation is completely removed from the data and as well the residuals maintain the feature of a white noise. Both ACF and PACF of the residual are basically zero for all lags, this is also an indication that, the model fit the data well.

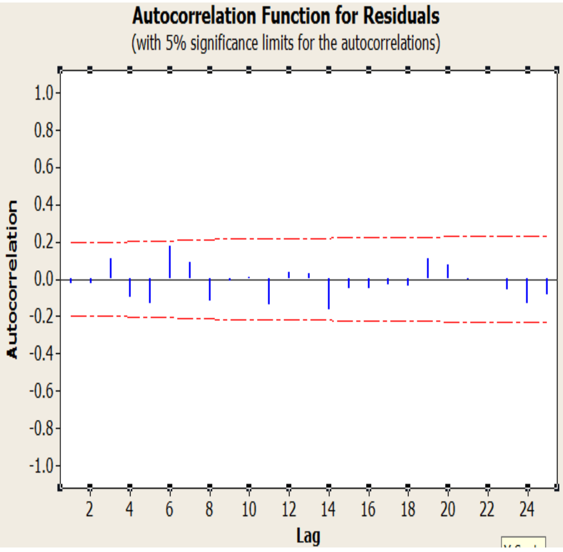

Fig. 6: ACF of the residuals

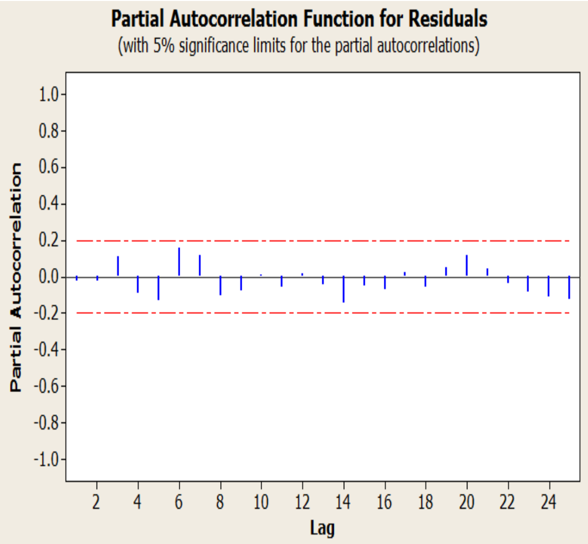

Fig. 7: PACF of the residuals

IV. Main result

Figures 8 and 9 show the comparisons between EWMA control chart for data and EWMA control chart for Residuals. Using equations (2 - 4), the original data on chemical product viscosity is plotted for EWMA chart using MINITAB 15 with a default value $\lambda = 0.2$.

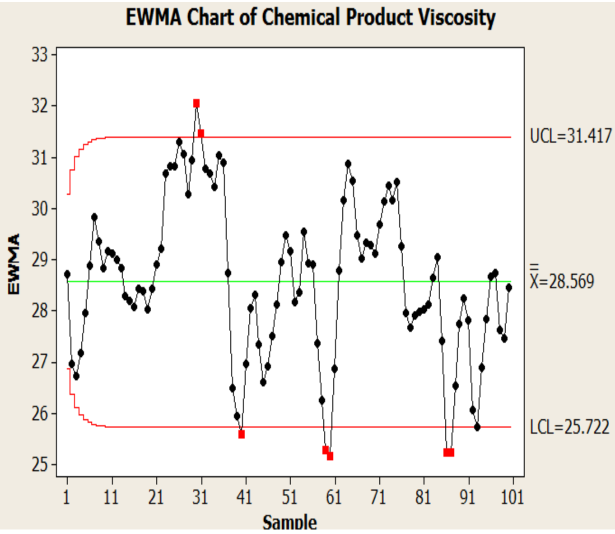

Fig. 8: EWMA Control Chart for data

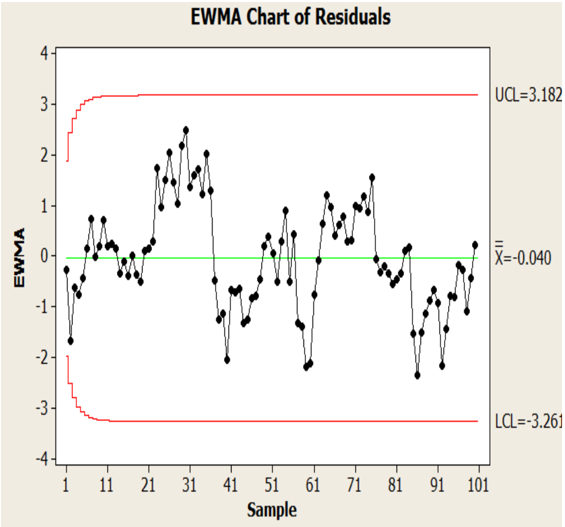

Fig. 9: EWMA Control Chart for Residual

It can easily be observed from Fig (8) that, some of the observations on EWMA control chart for the original data appear to be out of control due to some points shooting out beyond both the upper and lower control limits. Hence, the test failed at points 30, 31, 40, 59, 60, 86, 87. However, after modelling and confirmation of the removal of autocorrelation from the data, the EWMA control chart for residuals {see Fig (9)} indicates the processes are statistically in control.

V. Conclusion

In this study, we examined the effect of control chart performance when process data are autocorrelated. We presented modelling techniques that are aimed at removing autocorrelation effect from the process data using Box-Jenkins methodology. We also compared the performance of the methods before and after modelling the data. The results obtained reveal that, EWMA control charts can be used to detect a false alarm provided the process data exhibit the presence of autocorrelation. Furthermore, it is evident that, some popular control charts other than EWMA can also be employed to detect autocorrelation effect.

VI. Acknowledgement

The corresponding author would like to use this medium to offer his heartfelt appreciation and acknowledge the moral and financial support he received from his beloved mother in person of Hajiya Halima Abubakar Danbatta.

References

[1] Ajit Goswami, H.Nr. Dutta, Time Series Approaches to Statistical Process Control, *International Journal of Innovative Research in Computer Science and Technology,* 1(2), 33-37, 2013

[2] Akaike, H., A New Look at the Statistical Model Identification, *IEEE Transactions on Automatic Control*, 19(6), 716 -723, 1974

[3] Alwan, L. C., Roberts, H. V., Time Series Modeling for Statistical Process Control, *Journal of Business and Economic Statistics*, 6(1), 87-95, 1998

[4] Dobson, B., Control Charting Dependent Data, *Journal of Quality Engineering,* 7(4), 757-766, 1995.

[5] Duncan, A. J., *Quality Control and Industrial Statistics*, Irwin, 1986.

[6] Kandananond, K., The Detection of Shifts in Autocorrelated Process with MR and EWMA Charts, *The 7th Jordanian International Mechanical Engineering Conference*, Amman, Jordan, September 27th - 29th, 2010.

[7] Karaoglan, A. D., Bayhan, G. M., Performance Comparison of Residual Control Chart for Trend Stationary First Order AR Process, *Gazi University Journal of Science,* 24(2), 227-332, 2011.

[8] Lucass, J. M., Saccucci, M. S., Exponentially Weighted Moving Average Control Schemes: Properties and Enhancements, *Technometrics*, 32, 1-12, 1990.

[9] Montgomery and Johnson, Chemical Process Viscosity Data (12F6.3, robjhyndman.com/tsd/data/…/montgome… 2006).

[10] Montgomery, D. C., *Introduction to Statistical Quality Control*, John Wiley and Sons Inc, 2008.

[11] Montgomery, D. C., Runger, G. C., *Applied Statistics and Probability for Engineers*, John Wiley and Sons Inc, 2003.

[12] Nembhard, D. A., Nembhard, H. B., A Demerit Control Chart for Autocorrelated Data, *Quality Engineering*, 13(2), 179-190, 2000.

[13] Russo, S. L., Camargo, M. E., Fabris, J. P., *Application of Control Charts ARIMA for Autocorrelated Data*, 2012. DOI: 10.5772/50990.

[14] Sanders, N. R., Dan Reid, R., *Operations Management: An Integrated Approach*, John Wiley and Sons, 2002.

[15] Shey, H. S., Shin, L. L., Monitoring the Mean of Autocorrelated Observations with GWMA Control Chart Performance, *Journal of Statistical Computation and Simulation,* 79(12), 1393-1406, 2009.

[16] Shu, L. A., Tsung, F., Autcorrelated Process Monitoring Using Triggered Coscore Chart, *Quality and Reliability Engineering International Journal*, 18(5), 411-421, 2002.

[17] Steiner, S. H., *Quality Control and Improvement Based on Grouped Data*, Doctoral Dissertation, McMaster University Hamilton, Ontario, Canada, 1994.

[18] Woodal, W. H., Controversies and Contradictions in Statistical Process Control, *Journal of Quality Technology,* 27(1), 30-32, 2000.