SCHOOL OF STATISTICS
UNIVERSITY OF THE PHILIPPINES DILIMAN

WORKING PAPER SERIES

SPCR-Based Control Chart for Autocorrelated Processes with High Dimensional Exogenous Variables

by

Paul Eric G. Abeto
Joseph Ryan G. Lansangan

UPSS Working Paper No. 2016-07
November 2016

School of Statistics
Ramon Magsaysay Avenue
U.P. Diliman, Quezon City
Telefax: 928-08-81
Email: updstat@yahoo.com

Abstract

Monitoring processes in an industry is one means to ensure the quality of goods produced or services provided. Control charts are among the tools used to monitor processes. Control charts are constructed by estimating control limits wherein the process could be identified as stable. The estimation is made by analyzing the behavior of the monitored process. However, the assumptions of uncorrelatedness and normality of the measurements, common in most control charts, are sometimes uncharacteristic of the monitored process. Also, data from other variables may be available and may provide meaningful information on the behavior of the monitored process, and thus may be valuable in the estimation of the control limits. In this paper, a methodology of using sparse principal component regression from high dimensional exogenous variables to estimate control limits of autocorrelated processes is proposed. Simulations are made to further study different scenarios that may affect the proposed estimation. The false alarm rate, average run length during stable periods, and first detection rate upon structural change are used as key indicators for characterization and/or comparison. Simulation results suggest that modelling a process using high dimensional exogenous variables through sparse principal components creates better estimation of its corresponding control chart parameters. False alarm rates and average run lengths were comparable with the Exponentially Weighted Moving Average (EWMA) control chart. Also, faster identification of structural change was observed potentially due to the fact that the process is modelled in terms of other information carried by the exogenous variables.

Keywords: *Control chart, autocorrelated process, high dimensional data, sparse principal component regression*

1. Introduction

In monitoring activities or processes, it is critical to identify a group of performance indicators to assess if the activity/process is meeting the standards and expectations set by regulatory agencies, company targets, or consumer requirements. In the manufacturing industry, statistical process control (SPC) has long been used in order to monitor if the products manufactured conform to certain specifications or to regularly maintain the stability of the process to produce conforming products. Control charts are among the tools being used to monitor stability and conformity of the performance indicators identified by the company or possibly in general, a monitoring organization (Montgomery, 2009). Control charts are useful in identifying if the process is "in control," i.e. the process is stable, or if the process is "capable," i.e. the process meets the standards. There are several types of control charts, and commonly used are the Shewhart, cumulative sum or CUSUM, and EWMA, to name a few (Shewhart, 1931; Page, 1954; Roberts, 1959; Montgomery, 2009). The type of control charts varies depending on the characteristic of the data available, the objective of the control chart, and the objective of the organization in process monitoring.

Usually, creation and/or analysis of a control chart considers only the specific performance indicator being monitored. But "big data" or high dimensional data may be available, and a specific performance indicator may be affected by several exogenous factors. Information could be gathered by incorporating big data to identify possible contributors to the variability of a certain performance indicator. In the finance industry and stock brokerages for example, a monitored stock (the performance indicator in this case) could be related to the behavior of other different stocks, national and international indicators, or the activities of a website (e.g. views, trending topics, click-through rates, action-through rates). One of the key challenges in these scenarios is that a huge amount of data is available and considering all of these would not be economical or practical. Also, the input variables may exhibit high auto-correlations and cross-correlations which may affect the estimation process. Variable selection or in general dimension reduction and transformation, together with the appropriate estimation procedure or analytical approach, are thus critical.

While having big data might be beneficial in order to gather more information, certain key process indicators an organization is monitoring might exhibit autocorrelation, i.e. when an indicator is affected by its previous values. Autocorrelation may occur when the data gathering frequency is high or when a process has not really "stabilized" before obtaining a sample. Control charts are appropriate and work best when the process is stable or when the measurements are uncorrelated and normal, while autocorrelations and non-normality may induce errors in the analytical procedure. Such errors might lead to different effects such as detection of "instability" or out-of-control (OC) when there is really no shift in the process or vice versa, wherein the process is already OC but the control chart cannot detect such shifts (Maragah and Woodall, 1992; Montgomery, 2009).

This paper presents an estimation method for a univariate control chart model of an autocorrelated process that simultaneously incorporates inputs from high dimensional exogenous variables. In Section 2 the basic theory of control charts is given. Section 3 describes a dimension reduction technique integral to the proposed method. The proposed method is then presented in Section 4, and results of the simulation studies are in Section 5. Finally, concluding remarks and recommendations are presented in Section 6.

2. Control Charts

Shewhart Control Charts

Shewhart (1931) introduced the most basic and most widely used form of the control charts, frequently referred to as Shewhart charts. The use of such has several assumptions such as the normality of the data, independence of one data point from another, and stability of the data over time. Shewhart (1931) proposed the general theory of control charts for a specific variable of interest *w*, with mean μ_w and standard deviation σ_w, as follows (Montgomery 2009):

$$UCL = \mu_w + L\sigma_w$$
$$Centerline = \mu_w$$
$$LCL = \mu_w - L\sigma_w$$

The model proposed is a basic model akin to confidence intervals wherein L represents number of sigma units away from the mean. In the traditional use of control charts, L is usually equal to 3 and thus the level of significance is 0.0027. Also, the type of process parameter being monitored, as well as the number of data available, determines the type of control chart suitable for the process.

Shewhart control charts are useful for detecting large and sustained shifts in the process parameter, outliers, measurement errors, and data recording errors, among others, but it is not used for monitoring stable processes wherein normally only small shifts occur. In this regard, the cumulative sum or exponentially weighted moving average control charts are more appropriate for such task (Montgomery, 2009).

Exponentially Weighted Moving Average Control Chart

Exponentially weighted moving average (EWMA) charts are useful for detecting small shifts in the process. The control chart is by Roberts (1959) and its use expanded the field of reliability engineering. The estimation of the process parameter for EWMA is defined as:

$$z_i = \vartheta x_i + (1 - \vartheta) z_{i-1}$$

where $0 < \vartheta < 1$ and x_i are some constants and z_i the series of interest. EWMA considers the individual values as well as the immediate lagging value with a coefficient (or weight) equal to ϑ. If ϑ is equal to 1, it becomes similar to a normal Shewhart chart. As such using the said model, the control chart parameters are as follows (Montgomery, 2009):

$$UCL = \mu_0 + L\sigma \sqrt{\frac{\vartheta}{(2-\vartheta)}[1 - (1-\vartheta)^{2i}]}$$

$$Centerline = \mu_0$$

$$LCL = \mu_0 - L\sigma \sqrt{\frac{\vartheta}{(2-\vartheta)}[1 - (1-\vartheta)^{2i}]}$$

In addition, EWMA control charts are better for non-normal data when it comes to individual control charts since the Shewhart charts for individuals are more sensitive to non-normality of the parameter. Montgomery (2009) also indicated that "it is almost a perfectly nonparametric (distribution-free) procedure" given the robustness of the control chart to a wide range of applications.

Other control charts and process modeling

Runger and Willemain (1995) used nonoverlapping batch averages as the basic data and applied Shewhart charts when the process data is autocorrelated. Rather than getting the residuals from the correct time-series model of the autocorrelated process, as the residuals chart inherently degrades when autocorrelation increases, they presented a simplified derivation of the average run length through the weighted and unweighted batch means.

Mancenido and Barrios (2011) introduced the creation of control chart parameters using the AR-Sieve Bootstrap especially suited for autocorrelated process data. The method is better compared to traditional control chart models in terms of performance based on average run length for in-control and out-of-control processes. In addition, the advantage of creating models even for small samples sizes and distribution free parameters is met through the bootstrapping approach.

Poblador and Barrios (2015) extended the study by the introduction of exogenous variables in the model using transfer functions. The model incorporates the exogenous variables to better model the process parameter monitored. In comparison with EWMA, the study wields control limits that are narrower leading to higher false alarm rates but with higher detection rate for minimal structural changes.

Control Chart Performance Evaluation

As with other models and model generation techniques, there are parameters that evaluate the performance of the model generated. For control charts, since the tool is used to monitor the performance of a process parameter, the chart itself provides a near accurate assessment whether the process is in-control or out-of-control (OC).

Control chart models utilize the Average Run Length (ARL) in order to measure the number of data time points before an OC point is observed (Montgomery, 2009). In this regard, when the process is in-control, there should be no OC point, or with the ARL, it should be long enough not to detect "false alarms." In the same way, when the process is already out-of-control, the control chart should quickly detect process shifts, i.e., the ARL when a structural change occurs must be small. For Shewhart charts with control limits μ±3σ, this translates to an ARL for a stable process of about 371, which means that for the Shewhart chart, considering a stable process, it is expected that the interval between 2 false alarms is 371 time points.

The False Alarm Rate (FAR) is another indicator for control chart models. FAR is similar to a Type I error probability, measuring the proportion of instances where there are "false alarms" or points plotting out of the control limits when the process is in fact in-control (Besterfield, 2009). These points are detected during the stable period which should not have occurred. FAR is computed as the total number of out-of-control points under the stable phase. When the process is in-control, it is favorable that FAR would be as small as possible.

Equivalently, a control chart with faster detection rate after occurrence of a structural change would be more favorable. The FDR is computed as the length (or number of time points) until the first OC point after a process shift. In this regard, the first detection rate (FDR) should be as small as possible.

3. Dimension Reduction and Bootstrap Approaches

Sparse Principal Component Analysis

Principal component analysis (PCA) is a multivariate analysis technique in which the main and common objective is to reduce the dimensionality of a data set with correlated variables. PCA reduces the data through the construction of principal components, which are generally linear combinations of the original variables (Jolliffe, 2002). However, when dealing with nonstationary time series data, PCA induces the component loadings to be nearly the same for all variables (Barrios and Komoto, 2006; Lansangan and Barrios, 2009), which means that the interpretability and adequacy of using the principal scores may be compromised. To address the problem of interpretability, Lansangan and Barrios (2009) investigated the application of principal components to nonstationary time series data, and Lansangan (2013) further proposed using sparse principal components when dealing with high-dimensionality and nonstationarity in a regression problem. The works of Lansangan and Barrios (2009) and Lansangan (2013) are hinged on the framework introduced by Zou et al. (2006).

Zou et al. (2006) introduced a modification of PCA with an additional constraint and estimation procedure, such they called sparse principal component analysis (SPCA). They proposed an optimization problem to enhance the dimension reduction capabilities of PCA and obtain sparse loadings of the sparse principal components (SPCs). Using the SPCs allows for dimension reduction vis-à-vis interpretability. The SPCs, however, are correlated to some degree with each other, and explaining less variability than those of the PCs in PCA (Zou et al., 2006).

Model-based Bootstrap Method for Time Series

One approach for resampling for time series data is the model-based approach (Davison and Hinkley, 1997). In model-based resampling, there is a predefined model that is fitted into the data, and the residuals are then computed. Resamples are obtained from the generated residuals to generate new series. For a process $\{Y_t\}$ with an AR(1) time-series model, suppose $t = 1, 2, \dots, n$, then $Y_t = \phi Y_{t-1} + \varepsilon_t$. The autoregressive parameter is then estimated using conditional (general) least squares, so that the estimated innovations or the residuals take the form $\varepsilon_t = Y_t - \hat{\phi} Y_{t-1}$. Random samples (with replacement) are obtained from the generated residuals to obtain the simulated innovations $\varepsilon_0^*, \dots, \varepsilon_n^*$. From these the process is rebuilt, with $Y_0^* = \varepsilon_0^*$, and $Y_t^* = \phi Y_{t-1}^* + \varepsilon_t^*, t = 1, \dots, n$.

4. The SPCRAE Control Chart

A regression-based approach to constructing a control chart by modeling an autocorrelated output variable (i.e., the performance indicator) using high dimensional exogenous variables is presented. In summary, the control chart parameters are obtained through a sparse principal component regression with autocorrelated errors (**SPCRAE**). The steps are presented in Table 1.

Table 1. Methodology for the SPCRAE Control Chart parameters

STEP	DESCRIPTION	RESULT
Step 1	Generate the sparse principal components (SPCs) for the exogenous variables.	Sparse Principal Components (SPC_1, SPC_2, …, SPC_q)
Step 2	Fit the SPCs with the performance indicator, ***Y***, into a linear regression model with autocorrelated errors.	$y = \beta_0 + \beta_1 SPC_1 + \cdots + \beta_q SPC_q + \varepsilon$ $\varepsilon \sim AR(p)$
Step 3	Obtain the fitted values, $\hat{y}_t$. The fitted value at time t would serve as the centerline for the control chart at that time point.	$Centerline_t = \hat{y}_t$
Step 4	Obtain the residuals from *Step 2*.	$Residuals\ \{e_t\}$
Step 5	Generate a resample of the residuals $\{e_t\}$.	$Resample\ \{e_t^*\}$
Step 6	Repeat *Step 5* ***B*** times, where B is a large number, so that ***B*** independent replicates of reconstructed samples are generated.	$\{e_t^*\}_b, b \in \{1,2,\dots,B\}$
Step 7	Obtain the 0.00135th and 0.99865th percentiles for each resample $\{e_t^*\}_b$, $b \in \{1,2,\dots,B\}$	$\{e_t^*\}_b^{(0.00135)}, \{e_t^*\}_b^{(0.99865)}$
Step 8	Compute for the average of the 0.00135th and 0.99865th percentiles of the ***B*** resamples. $\widehat{\{e_t^*\}^{(0.99865)}} = \frac{\sum_1^B \{e_t^*\}_b^{(0.99865)}}{B}, \widehat{\{e_t^*\}^{(0.00135)}} = \frac{\sum_1^B \{e_t^*\}_b^{(0.00135)}}{B}$ These estimates would serve as the basis for the UCL and LCL of the process.	$UCL_t = \hat{y}_t + \widehat{\{e_t^*\}^{(0.99865)}}$ $LCL_t = \hat{y}_t + \widehat{\{e_t^*\}^{(0.00135)}}$

5. Simulation Study

Let the process of interest or output variable be y_t, and let the multiple exogenous p input variables be $x_{1t}, x_{2t}, \dots, x_{pt}$ across time points $t = 1, 2, \dots, n$. The process is assumed to be autocorrelated, while the input variables are high dimensional (the number of inputs p greater than the length or number of observations n) and elicit groupings within themselves. Table 2 summarizes the variable specifications.

Table 1. Variable Simulation settings

VARIABLE		SETTINGS
Input variables	Number of input variables, ***p***	Dependent on the series length. $p = 4n$
	Base variable parameters (for each latent group)	Group 1: $x_{grp1} \sim U[10{,}20]$ Group 2: $x_{grp2} \sim N(50, 8.333)$ Group 3: $x_{grp3} \sim U[100{,}110]$
	Non-base variable parameters	$x_i = \beta x_{base-grp} + \varepsilon$ $\varepsilon \sim N(0,1)$ → approx. 10% of base variable σ
Process / Output variable	Model generation	$y_t = y_{mt} + \varepsilon_t$
	y_{mt}	Average of all x's for that time period $y_{mt} = \frac{\sum_{i=1}^{p} x_p}{p}$
	Autocorrelation	$\varepsilon_t \sim AR(1)\ where\ \phi_{stable} = 0.6,\ v_t \sim N(0,1)$

To test the robustness and the behavior of the estimation procedure and/or the SPCRAE control charts, specific simulation scenarios are set as follows (Table 3).

Table 2. Scenario Parameter settings

SCENARIO PARAMETER	SETTINGS
Series Length	30 (Short)
	60 (Average)
	204 (Long)
Number of Latent Groupings (or Number of PCs)	2
	3
Structural Change in the Process	Mean Shift ($\hat{y}_{t_{shift}} = \hat{y}_t * 1.1$)
	Presence of trend (ϕ_{shift} = 0.9)
Number of SPCs used	Less than 1 of actual number of latent groupings, same as actual, more than 1 of actual

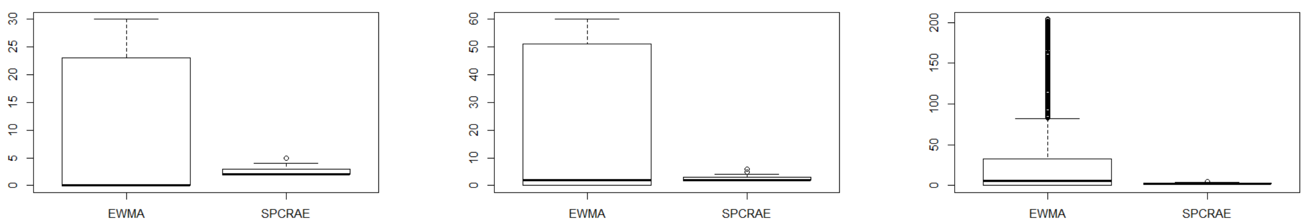
The simulation study is run with B=1000 bootstrap resamples for 1000 replicates per scenario. The SCPRAE control chart is compared to an EWMA control chart in terms of the 3 performance measures – false alarm rate (FAR), average run length (ARL) and first detection rate (FDR). The FAR is obtained by counting the number of OC points, i.e. points falling out of the control limits when the process is in the stable phase. The ARL is computed based from the average length of run between OC points for an in-control process, while the FDR is determined based from the first OC point after a process shift.

False Alarm Rate

False alarm rate is the measure of the number of out-of-control points even when the process is in control. For this measure, a smaller value of FAR is favorable as it indicates that the process would yield smaller chances of false detection.

Table 3 False Alarm Rate simulation results

Series Length	Number of PCs	Average		Percentage of replicates with smaller FAR		
		SPCRAE	EWMA	SPCRAE	EWMA	Tie
30	2	2.33	7.90	31.6%	66.5%	1.9%
30	3	2.30	9.05	46.2%	48.8%	5.0%
60	2	2.34	15.93	32.9%	64.5%	2.7%
60	3	2.35	18.70	62.6%	32.5%	4.9%
204	2	2.51	39.58	41.5%	51.8%	6.7%
204	3	2.52	51.57	71.9%	27.9%	0.2%
Total		**2.39**	**23.79**	**47.8%**	**48.7%**	**3.6%**



n=30 n=60 n=204

Figure 1. Boxplots of False Alarm Rates for EWMA and SPCRAE

From Table 3 False Alarm Rate simulation results

Series Length	Number of PCs	Average		Percentage of replicates with smaller FAR		
		SPCRAE	EWMA	SPCRAE	EWMA	Tie
30	2	2.33	7.90	31.6%	66.5%	1.9%
30	3	2.30	9.05	46.2%	48.8%	5.0%
60	2	2.34	15.93	32.9%	64.5%	2.7%
60	3	2.35	18.70	62.6%	32.5%	4.9%
204	2	2.51	39.58	41.5%	51.8%	6.7%
204	3	2.52	51.57	71.9%	27.9%	0.2%
Total		**2.39**	**23.79**	**47.8%**	**48.7%**	**3.6%**

, FAR is more consistent using SPCRAE than with EWMA, averaging at around 2.39 regardless of the scenario. It could also be noted that the average FAR for SPCRAE is lower than with that of EWMA (at 23.79). This can also be observed from the boxplots as shown in Figure 1. Also, there are almost the same number of replicates that SPCRAE is performing better than EWMA.

In more detail, Figure summarizes the different scenarios and combination of factors considered to check robustness and flexibility of SPCRAE and EWMA control charts with respect to FAR. It could be observed that using the SPCRAE methodology (Figure 2a), only the series length appears to affect the FAR, however, the FAR averages are small in terms of magnitudes. This implies that for either small or large number of observations, the methodology produces lower

FAR. For the two-way interactions between the scenario parameters, there are no significant patterns that could be observed for the identified scenarios. On the other hand, same can be observed about EWMA (Figure 2b), but as already noted, the values are larger than those of SPCRAE.

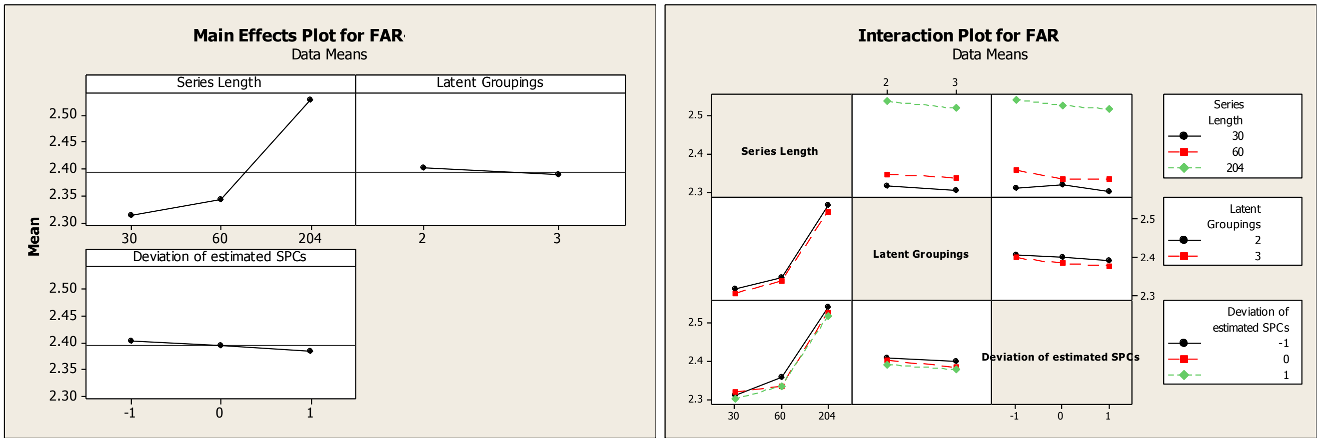


Figure 2a. SPCRAE

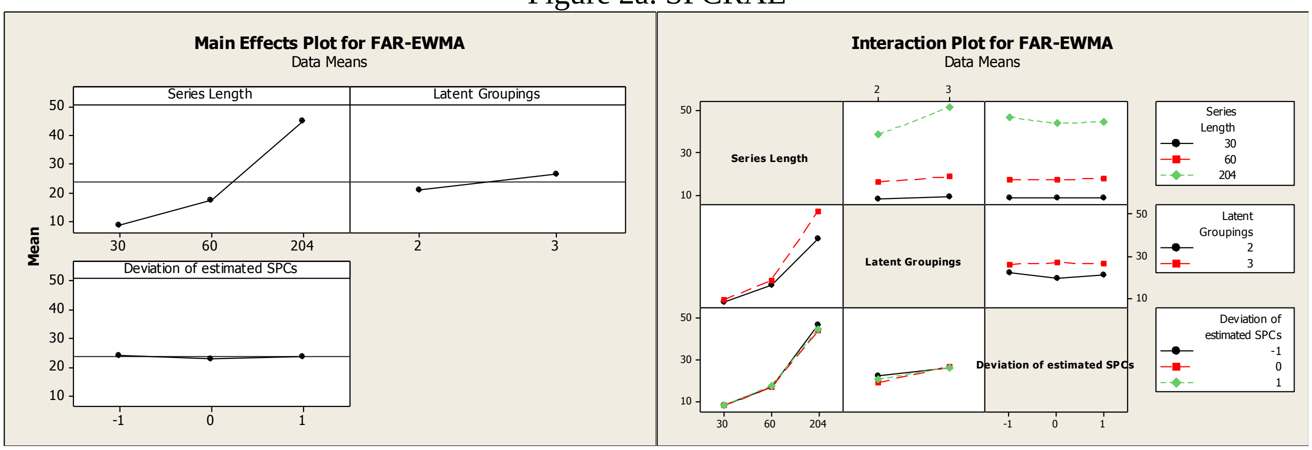


Figure 2b. EWMA

Figure 2. Effects charts for FAR of SPCRAE and EWMA Control charts

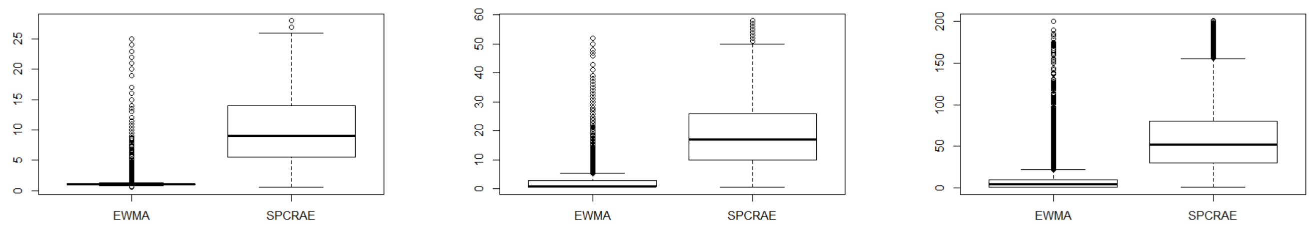
Average Run Length

The average run length (ARL) is the measure of length of runs in a given series between two OC points under an in-control phase. This measure is related to FAR as both consider the occurrence of OC points but this measure mainly identifies the length of time or the number of observations before the observer can expect the next OC point in the control chart after an OC point has been observed.

Table 4. Average Run Length simulation results

Series Length	**Number of PCs**	**Average**		**Percentage of replicates with longer ARL**		
		SPCRAE	**EWMA**	**SPCRAE**	**EWMA**	**Tie**
30	2	10.16	1.18	33.7%	65.5%	0.8%
30	3	10.15	1.73	52.3%	46.4%	1.4%
60	2	19.42	1.68	36.8%	62.9%	0.3%
60	3	19.67	3.49	66.4%	33.2%	0.4%

204	2	59.55	13.54	48.1%	51.8%	0.1%
204	3	60.11	6.25	70.2%	29.8%	0.0%
Total		**30.21**	**5.02**	**51.2%**	**48.3%**	**0.5%**



n=30 n=60 n=204

Figure 3. Boxplots of Average Run Length for EWMA and SPCRAE

On the average, the SPCRAE methodology produces longer runs than EWMA for in-control processes (see Table 5). In addition, looking at the comparative percentage of times ARL is longer, the SPCRAE methodology performs a bit better than EWMA. The boxplots per series length (Figure 3) suggest that SPCRAE consistently produces longer ARL.

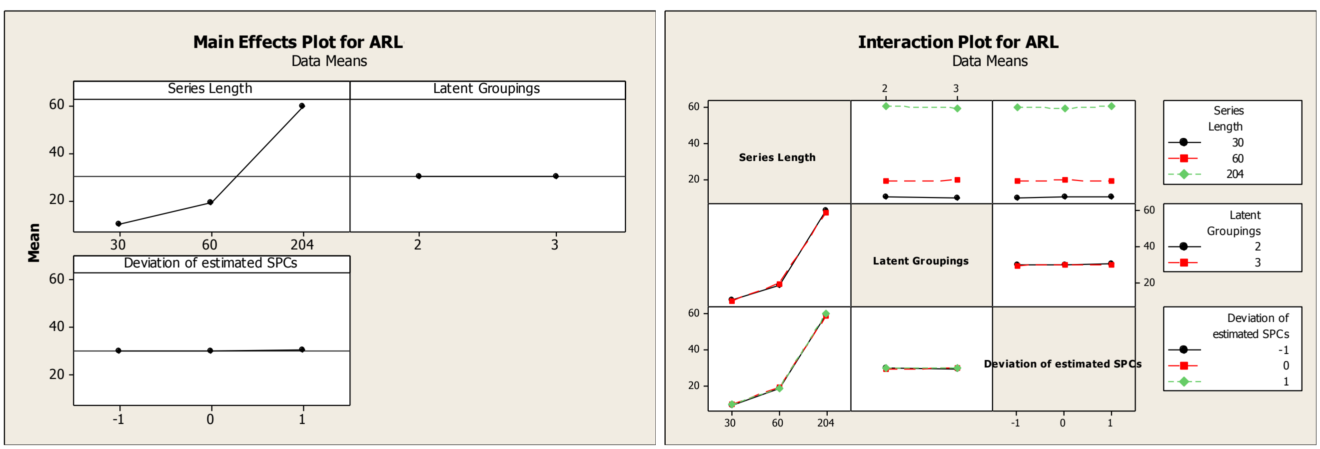


Figure 4a. SPCRAE

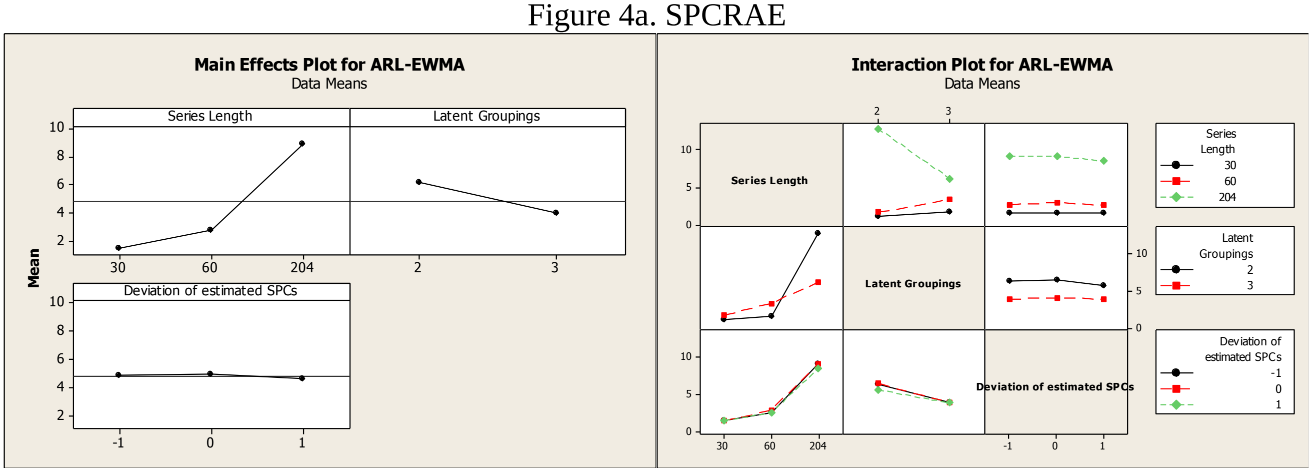


Figure 4b. EWMA

Figure 4. Effects charts for ARL of SPCRAE and EWMA Control charts

Generally speaking, SPCRAE produces longer run length for in-control processes than EWMA does. The ARL values of SPCRAE are consistently averaging at around 30% of the series length. From the percentage of replicates comparing both EWMA and SPCRAE, same with FAR, SPCRAE also performs consistently better when there are more inherent groupings among the variables under study.

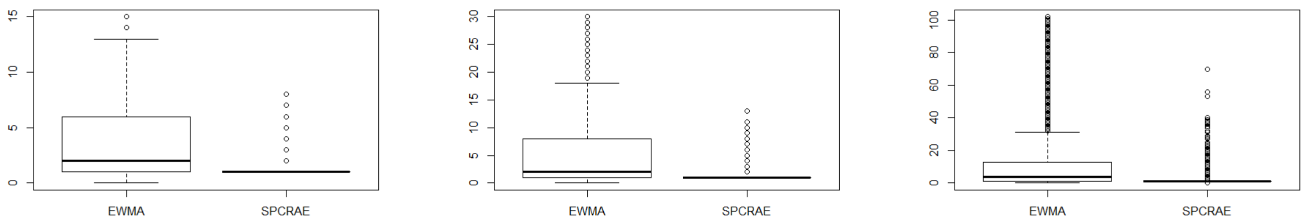
Figure 4 shows the ARL performances of SPCRAE and EWMA control charts with respect to the scenario parameters of the simulation study. From Figure , only the series length affect the ARL for the SPCRAE control charts. Looking at Figure , it could be observed that EWMA seems to perform more poorly when there are more latent groupings in the data. Also, ARL of EWMA seems to be affected by the interaction of series length and the number of latent groupings in such a way that when the series is longer, having more latent groupings tend to have shorter ARL.

First Detection Rate

The first detection rate (FDR) is the measure of the instance of detection of out-of-control after a structural change in the process. The measure would identify the response rate of the control chart in relation to the process shift. This is critical to the observer as this would measure how fast the system could adjust or identify unusual behavior.

Table 5 First Detection Rate simulation results

Series Length	**Number of PCs**	**Structural Change**	**Average**		**Percentage of replicates with smaller FDR**		
			SPCRAE	**EWMA**	**SPCRAE**	**EWMA**	**Tie**
30	2	Mean shift	1.24	2.00	19.4%	80.5%	0.1%
30	2	AR shift	1.01	2.39	53.3%	0.8%	45.9%
30	3	Mean shift	1.00	7.67	88.5%	11.0%	0.5%
30	3	AR shift	1.01	2.34	52.6%	0.9%	46.5%
60	2	Mean shift	1.48	8.23	48.3%	51.4%	0.3%
60	2	AR shift	1.01	2.03	43.7%	0.7%	55.6%
60	3	Mean shift	1.01	9.68	98.5%	0.3%	1.2%
60	3	AR shift	1.01	2.05	45.2%	0.4%	54.4%
204	2	Mean shift	3.53	31.64	93.3%	6.1%	0.6%
204	2	AR shift	1.01	1.98	42.5%	0.5%	57.0%
204	3	Mean shift	1.07	9.70	97.8%	0.5%	1.7%
204	3	AR shift	1.02	1.94	41.2%	0.6%	58.2%
Total			**1.28**	**6.80**	**60.4%**	**12.8%**	**26.8%**



n=30 n=60 n=204

Figure 5. Boxplot of First Detection Rate for EWMA and SPCRAE

The results shown in Table 5 First Detection Rate simulation results

Series Length	Number of PCs	Structural Change	Average		Percentage of replicates with smaller FDR		
			SPCRAE	**EWMA**	**SPCRAE**	**EWMA**	**Tie**
30	2	Mean shift	1.24	2.00	19.4%	80.5%	0.1%
30	2	AR shift	1.01	2.39	53.3%	0.8%	45.9%
30	3	Mean shift	1.00	7.67	88.5%	11.0%	0.5%
30	3	AR shift	1.01	2.34	52.6%	0.9%	46.5%
60	2	Mean shift	1.48	8.23	48.3%	51.4%	0.3%
60	2	AR shift	1.01	2.03	43.7%	0.7%	55.6%
60	3	Mean shift	1.01	9.68	98.5%	0.3%	1.2%
60	3	AR shift	1.01	2.05	45.2%	0.4%	54.4%
204	2	Mean shift	3.53	31.64	93.3%	6.1%	0.6%
204	2	AR shift	1.01	1.98	42.5%	0.5%	57.0%
204	3	Mean shift	1.07	9.70	97.8%	0.5%	1.7%
204	3	AR shift	1.02	1.94	41.2%	0.6%	58.2%
Total			**1.28**	**6.80**	**60.4%**	**12.8%**	**26.8%**

indicate that for all scenario parameter combinations, the SPCRAE methodology detects out-of-control faster than EWMA by about 5 observations on the average. In addition, the figures in Figure 5 indicate that after a structural change in the response variable, the process owner could expect to detect it within 1 to 2 observations or time points. This is much acceptable as in some cases, processes are monitored for longer intervals and thus the need for quicker detection of process shifts. Also, more than 85% of the time, SPCRAE performs better or at the same level as EWMA.

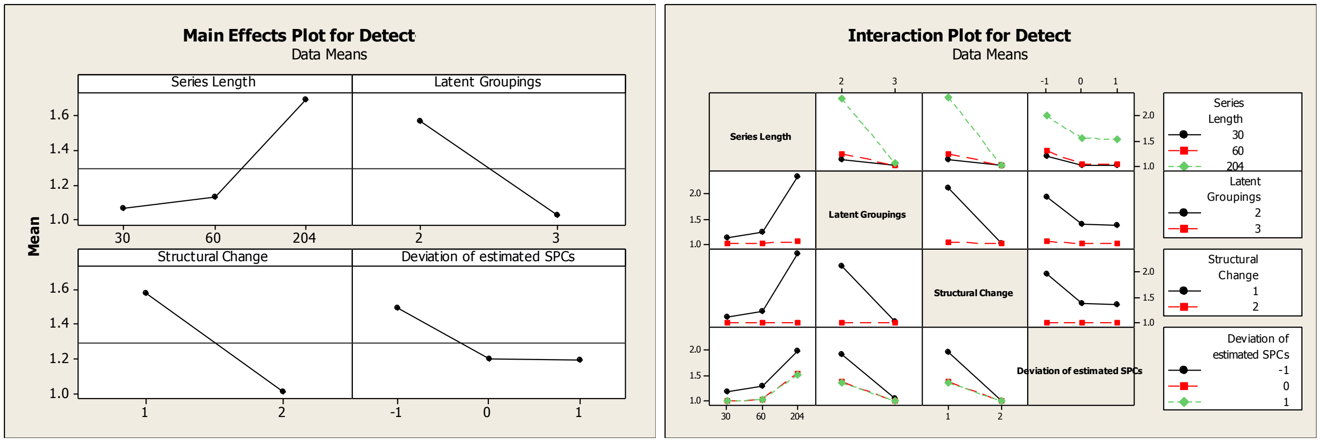


Figure 6a. SPCRAE

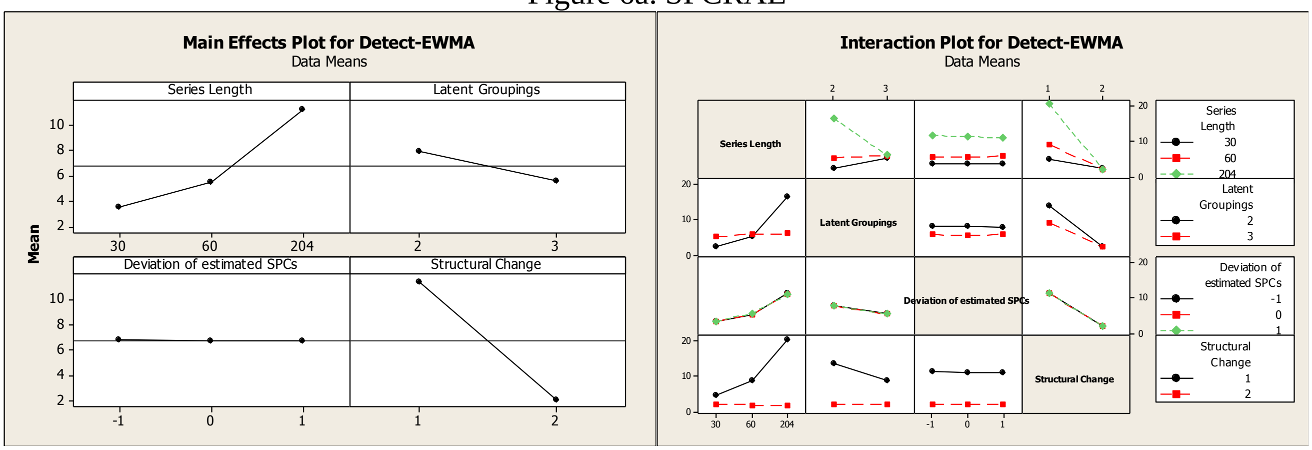


Figure 6b. EWMA

Figure 6. Effects charts for FDR of SPCRAE and EWMA Control charts

Looking at the simulation parameters that were changed, it could be observed that there may be significant contributors to the FDR (see Figure 6). In general, the series length, number of latent groupings, deviation of estimated SPCs, as well as type of structural change may affect FDR of SCPRAE and EWMA control charts.

6. Conclusions and Recommendations

This paper presents control limits estimation using the methodology of sparse principal component regression with autocorrelated errors (SPCRAE) to generate the SPCRAE control chart. Results show that SPCRAE control chart performs at least on the same level as an EWMA control chart in terms of average run length, false alarm rate, and first detection rate. Compared to EWMA, SPCRAE performs better with detection of out-of-control points after structural changes are introduced in the time series. The SPCRAE control charts are consistent with averages of about 2 to 3 FAR, at least 30% of the series length for ARL, and within 2 time points for FDR.

Clearly, modelling a process using high dimensional exogenous variables through sparse principal components is seen to produce better estimation of the process and its corresponding control chart parameters. Fewer false alarms, longer run lengths, and faster identification of structural change may be observed when much of the exogenous variables indeed are “in-favor” with or valuable to the process.

Recommendations for further studies include determining the effect of the amount of structural shifts in both the mean and the trend. The sensitivity of the control chart methodology with respect to process shifts might be more meaningful depending on the extent of the shift introduced in the process. The effect of the autocorrelated process could also be studied further in terms of processes other than an AR(1), or a more general model assumption for the process as a function of exogenous variables. In terms of high dimensionality, methods other than SPCR may be explored.

REFERENCES

BARNARD, G. A., 1959, Control Charts and Stochastic Processes, *Journal of the Royal Statistical Society Series B (Methodological)*, 21(2):239-271.

BARRIOS, E., and KOMOTO, K., 2006, Some Approaches in the Construction of a Sustainable Development Index for the Philippines, *International Journal of Sustainable Development & World Ecology*, 13(4):277-288.

BESTERFIELD, D.H. 2009, Quality Control, 8th ed., Upper Saddle River, New Jersey: Pearson Education Inc.

DAVISON, A.C. and HINKLEY, D.V., 1997, Bootstrap Methods and Their Application. Cambridge University Press.

JOLLIFFE, I.T. 2002. Principal Component Analysis, Second Edition. New York: Springer Verlag.

LANSANGAN, J.R.G., 2013, Sparse Principal Component Regression, *The Philippine Statistician*, 62(1):33-50.

LANSANGAN, J. and BARRIOS, E., 2009, Principal Components Analysis of nonstationary Time Series Data, *Statistics and Computing*, 19:173–187.

MANCENIDO, M. and BARRIOS, E., 2011, An AR-Sieve Bootstrap Control Chart for Autocorrelated Process Data, *Quality and Reliability Engineering International*, 28(4):387-395.

MARAGAH, H.D. and WOODALL, W.H., 1992, The Effect of Autocorrelation on the Retrospective X-chart, *Journal of Statistical Computation and Simulation*, 40:29-40.

MONTGOMERY, D.C., 2009, Introduction to Statistical Quality Control 6th edition. New York: Wiley.

PAGE, E.S., 1954, Continuous Inspection Schemes, *Biometrica*, 41(1-2):100-115.

POBLADOR, M.S.C. and BARRIOS, E., 2014, Control Chart for Monitoring Autocorrelated Process with Multiple Exogenous Inputs, *Forthcoming in Communications in Statistics – Simulation and Computation*.

ROBERTS, S.W., 1959, Control Chart Test Based on Geometric Moving Averages, *Technometrics*, 1(3):239-250.

RUNGER, G. C. and WILLEMAIN, T. R., 1995, Model-based and model-free control of autocorrelated processes, *Journal of Quality Technology*, 27(4):283-292.

SHEWHART, W.A., 1931, Economic Control of Quality of Manufactured Product, New York: D. Van Nostrand Company.

ZOU, H., HASTIE, T., and TIBSHIRANI, R., 2006, Sparse Principal Components Analysis, *Journal of Computational and Graphical Statistics*, 15(2): 265–286.