university of groningen / faculty of science and engineering

Overview of Techniques and Performance Comparison of Control Charts

Bachelor's Project: Mathematics

Student:	Li Voon Loke
First supervisor:	Dr. M. A. Grzegorczyk
Second assessor:	Dr. R. Jacobs

Department of Science and Engineering
UNIVERSITY OF GRONINGEN
Groningen, June 2019

Abstract

This paper gives an overview of different techniques used to construct a control chart, after which a comparison is made to compare their performance. The first section introduces the basic techniques, namely the Exponentially Weighted Moving Average (EWMA) and Cumulative Sum (CUSUM) control chart. These techniques are later used as a basis for monitoring a process modelled by a linear regression. The estimators of the regression model are obtained from ordinary least squares method and Bayesian inference. Results show that the Bayesian approach performs better on average under the EWMA control chart. A comparison between T^2 EWMA and T^2 CUSUM shows that the methods are similar and their performance depends on the values of the parameters. In general, smaller values of parameters in either control chart increases the sensitivity of the chart towards small shifts.

Contents

1

Introduction

Statistical process control (SPC) is a method of quality control which employs statistical methods to monitor a process. A popular SPC tool is the control chart, which is a graph used to study how a process changes over time. It was introduced by Walter A. Shewhart in the 1920's [8]. Traditionally, a process is monitored by collecting output samples and taking measurements that are associated with the quality characteristic that needs to be monitored, e.g. the thickness of contact lenses or concentration of certain chemicals in a drug, at specific time points. Currently, most control charts make use of statistical regression models between one or more response variables and a set of external process variables (covariates). The quality could then be checked by monitoring the functional relationship between the response and explanatory variables. During the production process it is monitored whether the model parameters stay unchanged over time, and in case of a parameter change, an alarm is given that the production process might be out of control and should be stopped. Hence, selecting the right control chart can reduce the number of false or missed alarms which is important in industries where upgrading machines, for example, is costly. Popular regression models that have been used for building control charts range from simple univariate to multiple multivariate linear regression models. Most of these models have been implemented in the standard way (frequentistic Statistics), but there were also models that took a Bayesian approach. The focus of this paper is to do an analysis and performance comparison between multiple types of control charts proposed in literature. Average run length (ARL), which is the average number of observations it takes for the chart to signal, is used in the measure of performance.
In this section, basic techniques (non-linear regression) used in constructing a control chart are described.

1.1 A Basic Design of Control Charts

SPC generally consists of two phases. The initial phase is called Phase I, in which an analysis is performed on existing (historical) data in order to determine the parameters for a stable, i.e. in-control process. When process parameters are unknown or unspecified, Phase I analysis must be done before process monitoring can begin in Phase II. Thus, most Phase II control charts assume that a reference sample is available from a corresponding Phase I analysis, from which the control limits can be estimated. Thus the success of prospective process monitoring in Phase II depends critically on the success of the Phase I analysis [6]. The focus of this paper is on

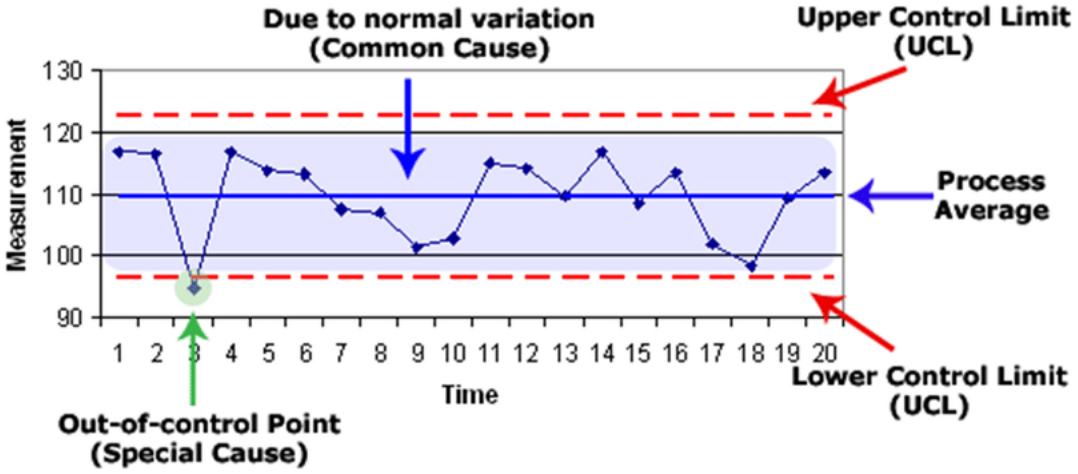

Figure 1.1: Components of a Control Chart, obtained from ResearchGate [4]

.

Phase II, so it is assumed that the in-control parameters are already known.
In Phase II, the quality characteristic of a process could be monitored using the sample mean $\bar{x}$. At each time period t, the average of n samples taken from the process is calculated. According to the Central Limit Theorem, the sample mean (when the process is in-control) is normally distributed with population mean μ and variance $\frac{\sigma^2}{n}$. So the lower and upper control limits (LCL and UCL respectively) could be set such that when the sample mean lies outside of the range, the process is likely to have gone out-of-control. The 99.73% confidence interval [16] can be found by setting UCL and LCL to:

$$LCL = \mu - 3\frac{\sigma}{\sqrt{n}} \tag{1.1}$$

$$UCL = \mu + 3\frac{\sigma}{\sqrt{n}} \tag{1.2}$$

The value 3 indicates the control limits 3σ away from the mean. Hence, this is also called the $\bar{X}$ or 3-*sigma* Shewart control chart. In practice, the value could be adjusted depending on how sensitive the control chart is to deviation from the target mean. However, the statistical disadvantages of the classic 3-*sigma* method are well-known [10][11], namely:

1. Standard deviation represents total variation in the data when control limits should represent only the random component.
2. It is not sensitive in detecting small shifts in the target parameter.
3. It is based on information about the process contained in the current observation only and ignores information given by the entire sequence of points.

There are several charting methods proposed for improvement. One example is the Exponentially Weighted Moving Average (EWMA) control chart which is the technique focused in this paper. A second method is called the Cumulative Sum (CUSUM) which will also be used in one of the comparisons in the last section.

1.2 Exponentially Weighted Moving Average (EWMA) Technique

The exponentially weighted moving average (EWMA) control chart was introduced by S.W. Roberts [19] in 1959, which is a good alternative to the Shewhart control chart when one is interested in small shifts. EWMA is a weighted average of all previous sample means. The formulation of the technique is as follows:

Take k samples, each of size n, let x_{ij} be the i^{th} measurement in the j^{th} sample. The mean of each sample is denoted

$$\bar{x_j} = \frac{\sum_{i=1}^{n} x_{ij}}{n}$$

The points z_j to be plotted on the chart depends on x_j:

$$\begin{aligned} z_j =& \lambda \bar{x_j} + (1-\lambda) z_{j-1} \\ =& \lambda \sum_{k=0}^{j-1} (1-\lambda)^k \bar{x}_{j-k} + (1-\lambda)^j z_0 \end{aligned} \tag{1.3}$$

where $z_0 = \mu_0$ is set to the target mean and $0 < \lambda \leq 1$ is the smoothing constant. The higher the value of λ, the more weight is given to the previous sample. The variance of z_i can be found, assuming that each measurement is independent:

$$\begin{aligned} \mathrm{Var}(z_j) =& \mathrm{Var}(\lambda \sum_{k=0}^{j-1} (1-\lambda)^k \bar{x}_{j-k} + (1-\lambda)^j z_0) \\ =& \lambda^2 \sum_{k=0}^{j-1} (1-\lambda)^{2k} \ \mathrm{Var}(\bar{x}_{j-k}) \\ =& \lambda^2 \sum_{k=0}^{j-1} (1-\lambda)^{2k} \frac{\sigma^2}{n} \\ =& \lambda^2 \frac{1-(1-\lambda)^{2j}}{1-(1-\lambda)^2} \frac{\sigma^2}{n} \\ =& \frac{\sigma^2}{n} \frac{\lambda}{2-\lambda} 1 - (1-\lambda)^{2j} \end{aligned}$$

Hence, the chart limits are given by:

$$\begin{aligned} LCL_j &= \mu_0 - L\sqrt{\mathrm{Var}(z_j)} \\ &= \mu_0 - L\sigma \sqrt{\frac{\lambda}{(2-\lambda)n}(1-(1-\lambda)^{2j})} \end{aligned} \tag{1.4}$$

$$UCL_j = \mu_0 + L\sigma \sqrt{\frac{\lambda}{(2-\lambda)n}(1-(1-\lambda)^{2j})} \tag{1.5}$$

where L is a multiplier that denotes the width of the control limits.

1.3 Cumulative Sum (CUSUM)

The CUSUM chart is proposed by Page in 1954 [18] that directly incorporates all the information in the sequence of sample values by plotting the cumulative sum of the deviations of the sample values from a target value [16]. If $\bar{x_j}$ is the average of the j^{th} sample and μ_0 is the target for the process mean, the cumulative sum of the control chart is

$$\begin{aligned} C_j &= \sum_{k=1}^{j} (\bar{x_k} - \mu_0) \\ &= (\bar{x_j} - \mu_0) + C_{j-1} \end{aligned} \tag{1.6}$$

Hence C_j is the cumulative sum of deviations up to and including the j^{th} sample. Similar to the EWMA chart, the CUSUM takes information from past observations which makes it more sensitive to small shifts in the process.

There are two plotting statistics in the CUSUM chart. They accumulate deviations that are above and below the target parameter respectively.

$$C_j^+ = max[0, x_j - (\mu_0 + K) + C_{j-1}^+] \tag{1.7}$$

$$C_j^- = max[0, (\mu_0 - K) - x_i + C_{j-1}^+] \tag{1.8}$$

where C_0^+=C_0^-=0

K is measured in units of standard deviation and depends on the size of the shift that needs to be detected. If either C_0^+ or C_0^- exceed the decision interval H, the process is considered out of control. The choice of H and K may be varied to obtain the desired in-control average run length (ARL) performance.

1.4 Average Run Length (ARL)

The Average Run Length (ARL) is defined as the average number of samples before the chart signals. ARL is generally employed as a performance indicator to evaluate the effectiveness of various control chart schemes, provided that the time interval between samplings remains constant [20]. When a process is in-control, longer ARL's indicates a lower false alarm rate. On the other hand, when a process is out-of-control, a shorter ARL indicates a better detection in process shifts. In this paper, the in-control ARL will be denoted ARL_0 and out-of-control ARL is denoted ARL_1. The average run length in this paper is obtained from the average of 10,000 simulations. The overall performance could also be computed from the Extra Quadratic Loss (EQL), which is the weighted average of $ARL_1's$ over the the entire domain of shifts (δ). It is defined as:

$$EQL = \frac{1}{\delta_{max} - \delta_{min}} \int_{\delta_{min}}^{\delta_{max}} \delta^2 ARL(\delta) d\delta$$

The EQL's are shown in Conclusion as a summary for the results.

2

Overview of Techniques

In the following sections, the EWMA control chart is combined with linear regression. The earliest reference appeared in a paper by DiPaola [9]. In 1969, Mandel [15] applied combined regression analysis and control chart theory to yield an effective technique for controlling man hours which is correlated with workload in the Washington Post Office Department.
In section 2.1, the relationship between the response and a co-variate is given only by a correlation value with the assumption that both the response and co-variate are normally distributed. It is shown later (Section 3.1) that the performance of the control chart is better when the process is monitored by its relationship with a co-variate compared to just monitoring the response alone. In Sections 2.2, the relationship between the response and co-variate is modelled by a linear regression while Section 2.3 applies the Bayesian approach to the linear regression. All 3 techniques are based on the EWMA control chart.

2.1 EWMA with a Single Auxiliary Variable

This method is described in the paper by Abbas, Riaz and Does [1]. Let the target response Y be correlated with a variable X, and the correlation between them is denoted ρ_{XY}. The observations of X and Y are obtained in the paired form for each sample. Assuming bi-variate normality of X and Y, i.e. $(X, Y) \sim N_2(\mu_Y, \mu_X, \sigma_Y^2, \sigma_X^2, \rho_{XY})$, the regression estimate is given by [5]

$$M_y = \bar{Y} + b_{XY}(\mu_X - \bar{X}) \tag{2.1}$$

where b_{XY} is the change in Y due to one unit change in X and $b_{XY} = \rho_{XY}\frac{\sigma_Y}{\sigma_X}$.
Using this in an EWMA control chart, the plotting statistic is

$$Z_j = \lambda M_{y_j} + (1-\lambda)Z_{j-1} \tag{2.2}$$

where λ is the smoothing constant and j denotes the j^{th} sample. The initial value Z_0 is set as the target mean, i.e. μ_Y. The control limits are

$$\begin{aligned} LCL_j &= \mu_Y - L\sigma_{M_Y}\sqrt{\frac{\lambda}{2-\lambda}(1-(1-\lambda)^{2j})} \\ UCL_j &= \mu_Y + L\sigma_{M_Y}\sqrt{\frac{\lambda}{2-\lambda}(1-(1-\lambda)^{2j})} \end{aligned} \tag{2.3}$$

where $\sigma_{M_Y}^2 = \frac{\sigma_Y^2 - b_{XY}^2\sigma_X^2}{n}$

It is also possible to monitor the process by just monitoring the average of each sample of Y alone. In this case, the plotting statistic is the same as that given in Section 1.2 where $\bar{x_j}$ in (1.3) is $\bar{Y_j}$. However, it can be shown (in the next chapter) that including the correlated variable in the control chart statistic leads to a lower ARL_1 when ARL_0 is kept constant.

2.2 EWMA Regression

Assume that the target response Y can be expressed as a linear function of a variable(s), i.e. $Y = AX + B +$ process variance. Kim, Mahmoud and Woodall [13] proposed the use of three univariate EWMA control charts to monitor the intercept, slope and variance of a process respectively. The motivation for their method was that if an assignable cause is present in a process characterised by a linear profile, then at least one of the parameters, i.e. the intercept, the slope or the variance would be affected. Hence, if at least one of the three charts gives an out-of-control signal, it means the process has gone out of control. Their method is based on the method by Kang and Albin [12].
Assume that when the process is in statistical control, the underlying model is

$$Y_{ij} = A_0 + A_1 X_i + \epsilon_{ij} \tag{2.4}$$

where j denotes the j^{th} sample and $i = 1, 2, ...n$ is the i^{th} observation in each sample. ϵ_{ij}'s are independent, identically distributed normal random variables with mean zero and variance σ^2.

Given a fixed set of values for X in each sample, X is transformed into X' such that $\bar{X}' = 0$. Then Equation (2.4) becomes:

$$Y_{ij} = B_0 + B_1 X_i' + \epsilon_{ij} \tag{2.5}$$

where $B_0 = A_0 + A_1\bar{X}$, $B_1 = A_1$, and $X_i' = (X_i - \bar{X})$.
The least squares estimators of A_0 and A_1 before the transformation have a bivariate normal distribution such that:

$$a_{0j} = \bar{y_j} - a_{1j}\bar{x} \ , \ a_{1j} = \frac{S_{xy(j)}}{S_{xx}}$$

where $S_{xy(j)} = \sum_{i=1}^{n}(x_i - \bar{x})y_{ij}$ and $S_{xx} = \sum_{i=1}^{n}(x_i - \bar{x})^2$
However, the least squares estimators of the transformed parameters are independent since the estimator of b_{0j} no longer contains the variable b_1:

$$b_{0j} = \bar{y_j} \ , \ b_{1j} = \frac{S_{xy(j)}}{S_{xx}}$$

So separate control charts for each parameters could be constructed without the problem of correlation by using model (2.5). The three plotting statistics using the EWMA control chart are then defined as follows:

$$\begin{aligned}
\text{Intercept}: &\quad Z_{Ij} = \lambda b_{0j} + (1-\lambda) Z_{I(j-1)} \\
\text{Slope}: &\quad Z_{Sj} = \lambda b_{1j} + (1-\lambda) Z_{S(j-1)} \\
\text{Variance}: &\quad Z_{Vj} = \max(\lambda \ \ln\ (MSE_j) + (1-\lambda) Z_{V(j-1)},\ \ln\ \sigma_0^2)
\end{aligned} \tag{2.6}$$

where $MSE_j = \frac{1}{n-2}\sum_i^n (y_{ij} - b_{1j}x_i' - b_{0j})$ is the usual OLS estimator of σ^2 based on residuals. Instead of monitoring the overall mean of each sample, the estimator of each parameter in the linear regression is monitored. The reason for using the log form of MSE is that an exact expression for $\text{Var}(\ln(MSE))$ has been derived before by Lawless [14]. However, in this paper, an approximation that is derived by Crowder and Hamiltion [7] is used. The control limits are as follows:

$$\begin{aligned}
\text{Intercept}: &\quad LCL_{Ij} = B_0 - L_I\sigma\sqrt{\frac{\lambda}{(2-\lambda)N}(1-(1-\lambda)^{2j})} \\
&\quad UCL_{Ij} = B_0 + L_I\sigma\sqrt{\frac{\lambda}{(2-\lambda)N}(1-(1-\lambda)^{2j})} \\
\text{Slope}: &\quad LCL_{Sj} = B_1 - L_S\sigma\sqrt{\frac{\lambda}{(2-\lambda)S_{xx}}(1-(1-\lambda)^{2j})} \\
&\quad UCL_{Sj} = B_1 + L_S\sigma\sqrt{\frac{\lambda}{(2-\lambda)S_{xx}}(1-(1-\lambda)^{2j})} \\
\text{Variance}: &\quad UCL_{Vj} = \ \ln\ \sigma_0^2 + L_V\sqrt{\frac{\lambda}{2-\lambda}\text{Var}(\ln(MSE_j))\ (1-(1-\lambda)^{2j})}
\end{aligned} \tag{2.7}$$

where $\text{Var}(\ln(MSE_j)) \approx \frac{2}{n-2} + \frac{2}{(n-2)^2} + \frac{4}{3(n-2)^3} - \frac{16}{15(n-2)^5}$, N is the number of samples, L_I, L_S, L_V are the width of the corresponding chart and could be adjusted to obtain the desired ARL_0. The variance is monitored by only the Upper Control Limit because in most processes, we are only interested in the increase in variance.

2.3 Bayesian Approach

Abbas et. al. [2] proposed a Bayesian alternative to the technique described above by Kim, Mahmoud and Woodall. The technique is mostly the same, except that the estimators of B_0, B_1 are obtained from the posterior distribution of the parameters. According to Bayes theorem:

$$P(\gamma|D) = \frac{L(D|\gamma)P(\gamma)}{P(D)} \tag{2.8}$$

where D and γ denote the data and the parameter respectively, $P(\gamma|D)$ is the posterior distribution, $L(D|\gamma)$ is the likelihood function, $P(\gamma)$ is the prior distribution of the parameter, and $P(D)$ is the distribution function of the data. Since the observed data D is fixed, Equation (2.8) can also be written as:

$$P(\gamma|D) \propto L(D|\gamma)P(\gamma)$$

Using the same regression model of Equation (2.5) and assuming $L(Y|B_0, B_1, \sigma) \sim N(B_0 + B_1X', \sigma^2)$, $B_0 \sim N(b_0, \kappa_0^2)$ and $B_1 \sim N(b_1, \kappa_1^2)$, the joint posterior distribution of B_0 and B_1 is given as:

$$\begin{aligned} P(B_0, B_1|Y) \propto & \exp(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_{ij} - B_0 - B_1X_i')^2) \\ & \times \exp(-\frac{1}{2\kappa_0^2}(B_0 - b_0)^2) \\ & \times \exp(-\frac{1}{2\kappa_1^2}(B_1 - b_1)^2) \end{aligned} \tag{2.9}$$

Since B_0 and B_1 are independent as mentioned previously, the joint distribution is equal to the product of the individual marginal distribution. After expanding and simplifying the expression in (2.9) [see Appendix A.1 for derivation], we get $B_0 \sim N(b_{0nj}, \sigma_{0nj}^2)$ and $B_1 \sim N(b_{1nj}, \sigma_{1nj}^2)$ where the Bayes estimators of B_0 and B_1 are:

$$b_{0nj} = \frac{n\bar{y_j}\kappa_0^2 + b_0\sigma^2}{n\kappa_0^2 + \sigma^2}, \qquad b_{1nj} = \frac{n\bar{y_j}\kappa_1^2 + b_1\sigma^2}{n\kappa_1^2 + \sigma^2} \tag{2.10}$$

The Bayes estimators of the variances are:

$$\sigma_{0nj}^2 = \frac{\kappa_0^2\sigma^2}{n\kappa_0^2 + \sigma^2}, \qquad \sigma_{1nj}^2 = \frac{\kappa_1^2\sigma^2}{n\kappa_1^2 + \sigma^2} \tag{2.11}$$

Finally, taking the prior of the variance σ^2 as inverse gamma distribution, $\sigma^2 \sim IG(\alpha_0, \beta_0)$, the posterior distribution is:

$$\begin{aligned} P(\sigma^2|Y) \propto & (\sigma^2)^{-\frac{n}{2}} \exp(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_{ij} - B_0 - B_1X_i')^2) \\ & \times (\sigma^2)^{-\alpha_0 - 1} \exp(-\frac{1}{\sigma^2}\beta_0) \end{aligned} \tag{2.12}$$

Again, after expanding and simplifying, the posterior distribution is $\sigma^2 \sim IG(\alpha_{nj}, \beta_{nj})$ where the Bayes estimators are:

$$\alpha_{nj} = \alpha_0 + \frac{n}{2}, \qquad \beta_{nj} = \beta_0 + \frac{(n-2)\sigma_0^2}{2} \tag{2.13}$$

and $\sigma_0^2 = \sum_{i=1}^{n}(y_{ij} - B_0 - B_1\bar{X}_i')^2$. The expectation of σ^2, $E(\sigma^2) = \frac{\beta_{nj}}{\alpha_{nj}-1}$ is then used in Equations (2.10) and (2.11) in place of σ^2.

From Equations (2.10) and (2.11), the plotting statistics are obtained. Hence the three univariate Bayesian control charts for monitoring the Y-intercept (B_0), slope (B_1) and variance (σ^2) are defined respectively as follows:

$$\begin{aligned}
\text{Intercept}: &\quad Z_{Ij} = \lambda b_{0nj} + (1-\lambda) Z_{I(j-1)} \\
\text{Slope}: &\quad Z_{Sj} = \lambda b_{1nj} + (1-\lambda) Z_{S(j-1)} \\
\text{Variance}: &\quad Z_{Vj} = \max(\lambda \ \ln\ (E(\sigma^2)) + (1-\lambda) Z_{V(j-1)},\ \ln\ \sigma_0^2)
\end{aligned} \tag{2.14}$$

where $0 < \lambda \leq 1$ is the smoothing constant, $Z_{I0} = B_0$, $Z_{S0} = B_1$, $Z_{V0} = \ln\ (1)$. The corresponding control limits are then defined as follows:

$$\begin{aligned}
\text{Intercept}: &\quad LCL_{Ij} = B_0 - L_I \sigma_{onj} \sqrt{\frac{\lambda}{(2-\lambda)}(1-(1-\lambda)^{2j})} \\
&\quad UCL_{Ij} = B_0 + L_I \sigma_{onj} \sqrt{\frac{\lambda}{(2-\lambda)}(1-(1-\lambda)^{2j})} \\
\text{Slope}: &\quad LCL_{Sj} = B_1 - L_S \sigma_{1nj} \sqrt{\frac{\lambda}{(2-\lambda)}(1-(1-\lambda)^{2j})} \\
&\quad UCL_{Sj} = B_1 + L_S \sigma_{1nj} \sqrt{\frac{\lambda}{(2-\lambda)}(1-(1-\lambda)^{2j})} \\
\text{Variance}: &\quad UCL_{Vj} = \ln\ (1) + L_V \sqrt{\frac{\lambda}{2-\lambda}\ \mathrm{Var}(ln(E(\sigma^2)))}
\end{aligned} \tag{2.15}$$

2.4 T^2 EWMA and T^2 CUSUM Regression

Finally, in this section, another version (generally known as the T^2 statistic) of the non-Bayesian EWMA regression is described, and compared against non-Bayesian CUSUM method.

Up to this point the parameters in a linear regression are tested individually. Instead of constructing three control charts, the parameters of a linear regression could be collected in a matrix and combined into one plotting statistic. In the model used in this simulation, i.e. $\boldsymbol{Y_k} = 13 + 2\boldsymbol{X'} + \boldsymbol{\epsilon}$ described above, the k^{th} sample of the response Y is an $n \times 1$ vector, X is an $n \times 2$ design matrix, the parameters are stored in a 2×1 vector $\boldsymbol{\beta}$, and $\boldsymbol{\epsilon}$ is an $n \times 1$ vector, where $n = 4$ is the number of observations in a sample. The relationship can be represented as:

$$\begin{bmatrix} Y_{1k} \\ Y_{2k} \\ Y_{3k} \\ Y_{4k} \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \\ 1 & x_4 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} \epsilon_{1k} \\ \epsilon_{2k} \\ \epsilon_{3k} \\ \epsilon_{4k} \end{bmatrix} \tag{2.16}$$

β_0 and β_1 are known in this model and their values are 13 and 2 respectively. The least squares estimate of $\boldsymbol{\beta}$ is denoted:

$$\hat{\boldsymbol{\beta}}_{\boldsymbol{k}} = (\mathbf{X^T X})^{-1} \mathbf{X^T Y_k}$$

The EWMA statistic of the k^{th} sample is a vector, given in the paper by Noorossana et. al. [17]:

$$\boldsymbol{z_k} = \lambda(\hat{\boldsymbol{\beta}}_{\boldsymbol{k}} - \boldsymbol{\beta})^T + (1-\lambda)\boldsymbol{z_{k-1}}$$

and the plotting statistic is a single value denoted:

$$T^2_{z_k} = \boldsymbol{z_k}\Sigma^{-1}\boldsymbol{z_k}^T$$

where Σ^{-1} is the inverse covariance matrix of $\hat{\boldsymbol{\beta}}_{\boldsymbol{k}}$. The elements of Σ are given in Appendix A.2 while the proof is given in the paper by Noorossana et.al. Since $T^2_{z_k}$ is always positive, the control limit of this chart is only a single value h. Hence, the chart signals when $T^2_{z_k} > h$. Again, h can be set to achieve the desired ARL_0, which in this case would be approximately 300.

This paper focuses heavily on the EWMA charting technique. However, it might also be useful to mention how it compares to the CUSUM method of the same set up. Since the target values are now a vector, the method described in Section 1.3 has to be modified accordingly such that the maximum of a vector and the null vector could be interpreted. A full explanation is given in Section 2, page 292 by Crosier [3]. In summary, the CUSUM T^2 statistic is expressed as follows:

$$s_k = [(C_{k-1} + \hat{\beta}_k - \beta)^T\Sigma^{-1}(C_{k-1} + \hat{\beta}_k - \beta)]^{1/2}$$

$$C_k = \begin{cases} (C_{k-1} + \hat{\beta}_k - \beta)(1 - \frac{K}{C_j}), & \text{if } s_k > K \\ \mathbf{0}, & \text{otherwise} \end{cases}$$

where $C_0 = \mathbf{0}$ and $K > 0$. So the chart statistic is:

$$T^2_{C_k} = C_k^T\Sigma^{-1}C_k$$

2.5 Derivation of Control Limits

Most literature do not give an explicit explanation on how to derive the width of the control limits, i.e. the values of L_I, L_S, L_V, and h. Hence, in this section, the method used to approximate these values is given.

As mentioned previously, the width of the limits are set according to the desired ARL_0. Assume that the desired $ARL_0 = 200$, which means the probability of obtaining a false alarm is $\frac{1}{200} = 0.005$. Hence, the control limits should be set such that 99.5% of the in-control samples should fall within the range. Theoretically, this could be done by generating a large number of in-control samples and finding the corresponding percentiles. Since the control limits for monitoring the slope and intercept are two-sided, the values between the lower and upper percentiles should constitute 99.5% of the total samples. Hence, the lower percentile is 0.025 and the upper percentile is 0.975 as illustrated in Figure 2.1. However, after generating 10,000 samples and obtaining the estimates of of L_I, L_S and L_V, it is found that the resulting ARL_0 is not precise enough as the chart is rather sensitive to the values of L up to 2 decimal places. Hence, the width of the control charts is manually adjusted by increasing or decreasing the values until the desired ARL_0 is achieved.

In the next sections, a performance comparison and analysis using simulated data will be performed on the techniques described above.

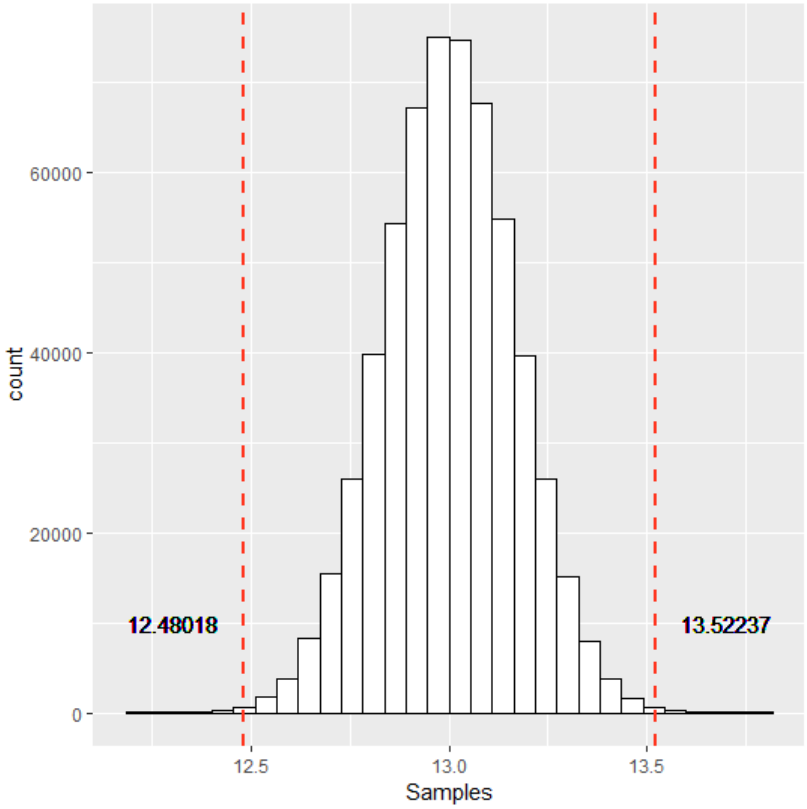

Figure 2.1: Upper and Lower Percentiles of the Samples that Center at 13

3

Simulation Study and Results

3.1 EWMA with a Single Auxiliary Variable

Here the technique described in Section 2.1 is compared with the classic EWMA control chart (without an auxiliary variable). Let Y and X both be generated from the standard normal distribution. Fixing the smoothing constant λ, width of control limits L, out-of-control shift δ, and in-control ARL_0, the correlation between Y and X are varied. [See Appendix A.3 for code.]

Table 3.1: ARL_1 values for $\lambda = 0.1$, $\delta = 0.5$, $L = 2.824$, $ARL_0 = 500$

Correlation	With Auxiliary Variable	Classical EWMA
0.05	8.0915	8.1092
0.25	7.8378	8.1580
0.50	6.7668	8.1393
0.75	4.7454	8.0947
0.95	2.1312	8.2120

As shown from the table, if the target response is correlated with an auxiliary variable, using the statistic that includes the variable shortens the ARL_1 with increasing correlation while the ARL_1 does not change for the traditional EWMA that does not include the variable, which is expected. This provides evidence to support that a target quality should be monitored as a linear profile.

3.2 Bayesian vs. Non-Bayesian (EWMA)

Assuming that the underlying model is $Y = 13 + 2X' + \epsilon$ and X' takes the values $(-3, -1, 1, 3)$ such that $\bar{X}' = 0$, 100 in-control samples are simulated. The process is assumed to have a natural variance of 1, so $\epsilon \sim N(0, 1)$. A shift (indicating out-of-control situation) is introduced at the 101^{th} sample and the simulation continues until the control chart detects the shift (i.e. when a sample crosses the control limits). The out-of-control samples are divided into 3 cases, i.e. a shift in the intercept, a shift in the slope and a shift in the process variance. Furthermore, the degree of shifts (δ, in units of standard deviation) and the value of the smoothing constant λ are varied to study how the ARL_1 changes. Lastly, the value of the control limits L_I, L_S, L_V where $I =$ Intercept, $S =$ Slope, $V =$ Variance are set such that the ARL_0 for all cases remain approximately the same for a fair comparison. In the

simulation study presented in the following sections, each combined control chart is set to have $ARL_0 \approx 300$. [See Appendix A.4 for code.] The results are shown in Table 3.2, which includes the effect of varying values of smoothing constant λ. Further illustration is given in Section 3.4 to graphically demonstrate its effect. The results show that the smaller the value of λ, the more sensitive it is towards small shifts.

For the Bayesian approach, the hyper-parameters b_0, b_1, κ_0^2, κ_1^2, α_0, β_0 need to first be determined. In Section 2.7 of the paper by Abbas et. al. [2], the technique used to elicitate hyper-parameters is described which involved consulting expert opinions about the parameters under study. This paper does not study in-depth such technique. However, the simulations are conducted under various values of hyper-parameters to observe their effect on the performance of the control charts. The results are shown in Table 3.3-3.5. For a shift in Y-intercept (Table 3.3), there is a continuous decrease in ARL_1 when the location hyper-parameters increase. On the other hand, the ARL_1 decreases slightly with a decrease in scale hyper-parameters while there is not much difference in performance for the change of hyper-parameters for variance. For a shift in slope and process variance (Table 3.4 & Table 3.5), there is no significant trend in the performance for varying values of hyper-parameters. A comparison with the non-Bayesian approach is shown in Table 3.6, where $\lambda = 0.2$, $b_0 = 20$, $b_1 = 2.5$, $\kappa_0^2 = 35$, $\kappa_1^2 = 12$, $\alpha_0 = 0.68$, $\beta_0 = 0.3$. The results show that the Bayesian approach on average performs better than the non-Bayesian approach.

3.3 T^2 EWMA vs. T^2 CUSUM Regression

The regression model remains the same as described above, substituting sample values into the equations described in Section 2.4, the result of T^2 EWMA control chart is shown in Table 3.7. The ARL_1 values in general decreases very quickly even for small shifts. Comparing it to monitoring the estimators separately with three control charts, the T^2 method performs better for shifts in intercept and slope. Shifts in variance are not detected as quickly when the shift is smaller than 2 times the in-control variance. The negative values are due to the fact that in most runs of the simulation, the shift is detected immediately as it happens, so combined with certain runs that give false alarms even before the shift has happened, the average run length goes to negative. For example, the first 100 samples are in-control, so if 9 out of 10 runs detect the shift at the 101^{th} sample and 1 run gives a false alarm at the 90^{th} sample, the average run length is $(9 \cdot 101 + 1 \cdot 90)/10 - 100 = -0.1$. Since it does not make sense to have a negative ARL, the negative values should be interpreted as immediate detection of the shifts. Table 3.8 shows the result for T^2 CUSUM chart when the parameter $K = 0.5$, while Table 3.9 is the result for $K = 1$. The CUSUM chart performs better when $K = 0.5$ and worse when $K = 1$ compared to the EWMA chart with $\lambda = 0.2$.

Table 3.2: Non-bayesian ARL_1 values for different shifts δ and smoothing constants λ

δ_I	Intercept Shifts ($B_0 + \delta_I\sigma$)		
	$\lambda = 0.05$	$\lambda = 0.2$	$\lambda = 0.5$
0.0	307.20	293.11	303.11
0.2	36.83	74.98	148.41
0.4	12.16	18.47	40.25
0.6	6.75	8.67	14.18
0.8	4.59	5.53	7.33
1.0	3.58	4.14	4.82
1.2	3.01	3.36	3.66
1.4	2.61	2.88	3.04
1.6	2.38	2.57	2.65
1.8	2.22	2.35	2.41
2.0	2.11	2.20	2.23
δ_S	Slope Shifts ($B_1 + \delta_S\sigma$)		
	$\lambda = 0.05$	$\lambda = 0.2$	$\lambda = 0.5$
0.000	300.27	291.27	299.13
0.025	195.48	253.50	278.66
0.050	92.89	168.75	234.18
0.075	49.08	101.82	179.42
0.100	30.66	60.15	125.91
0.125	21.26	38.38	86.71
0.150	16.11	26.40	60.20
0.175	12.63	19.22	42.39
0.200	10.41	14.54	30.11
0.225	8.74	11.72	22.65
0.250	7.45	9.65	17.01
δ_V	Variance Shifts ($\delta_V\sigma$)		
	$\lambda = 0.05$	$\lambda = 0.2$	$\lambda = 0.5$
1.0	302.56	295.25	309.58
1.2	37.62	40.34	42.97
1.4	10.14	13.03	13.89
1.6	5.35	6.98	7.24
1.8	3.77	4.77	4.99
2.0	3.09	3.69	3.82
2.2	2.74	3.19	3.25
2.4	2.50	2.82	2.90
2.6	2.36	2.61	2.66
2.8	2.27	2.46	2.50
3.0	2.21	2.35	2.40

Table 3.3: Bayesian ARL_1 values for shifts in Y-intercept at $ARL_0 = 300$

δ_I	Sensitivity Analysis of Hyper-parameters (Location)			
	$b_0 = 15$, $b_1 = 1.5$	$b_0 = 20$, $b_1 = 2.5$	$b_0 = 25$, $b_1 = 3.5$	$b_0 = 30$, $b_1 = 4.5$
0.0	303.14	303.25	296.93	298.72
0.2	65.64	48.83	40.85	38.20
0.4	16.87	14.52	13.23	13.00
0.6	8.22	7.50	7.21	7.25
0.8	5.38	5.03	4.92	4.93
1.0	4.05	3.83	3.78	3.80
1.2	3.30	3.17	3.13	3.13
1.4	2.85	2.75	2.72	2.71
1.6	2.55	2.46	2.44	2.45
1.8	2.34	2.28	2.26	2.72
2.0	2.19	2.16	2.14	2.14
	Sensitivity Analysis of Hyper-parameters (Scale)			
	$\kappa_0^2 = 45$, $\kappa_1^2 = 16$	$\kappa_0^2 = 35$, $\kappa_1^2 = 12$	$\kappa_0^2 = 25$, $\kappa_1^2 = 8$	$\kappa_0^2 = 20$, $\kappa_1^2 = 6$
0.0	309.87	310.16	301.70	303.47
0.2	62.21	61.27	57.81	54.82
0.4	15.72	15.60	15.19	14.59
0.6	7.45	7.25	7.25	7.14
0.8	4.72	4.67	4.64	4.82
1.0	3.40	3.35	3.35	3.44
1.2	2.73	2.71	2.68	2.73
1.4	2.31	2.28	2.27	2.32
1.6	1.99	1.97	1.95	2.01
1.8	1.76	1.75	1.77	1.78
2.0	1.63	1.62	1.62	1.66
	Sensitivity Analysis of Hyper-parameters (Variance)			
	$\alpha_0 = 0.1$, $\beta_0 = 0.1$	$\alpha_0 = 0.68$, $\beta_0 = 0.3$	$\alpha_0 = 1.5$, $\beta_0 = 0.9$	$\alpha_0 = 2.5$, $\beta_0 = 2.0$
0.0	311.27	299.89	301.53	298.15
0.2	62.28	64.92	67.03	66.34
0.4	15.72	16.13	16.43	16.26
0.6	7.44	7.63	7.49	7.62
0.8	4.73	4.74	8.77	4.72
1.0	3.41	3.43	3.40	3.42
1.2	2.72	2.69	2.72	2.70
1.4	2.30	2.29	2.29	2.28
1.6	2.00	1.98	1.98	1.98
1.8	1.74	1.75	1.72	1.71
2.0	1.62	1.62	1.57	1.58

Table 3.4: Bayesian ARL_1 values for shifts in slope at $ARL_0 = 300$

δ_S	Sensitivity Analysis of Hyper-parameters (Location)			
	$b_0 = 15$, $b_1 = 1.5$	$b_0 = 20$, $b_1 = 2.5$	$b_0 = 25$, $b_1 = 3.5$	$b_0 = 30$, $b_1 = 4.5$
0.000	306.58	298.85	302.19	301.49
0.025	269.23	252.16	263.34	277.01
0.050	180.47	178.46	202.71	227.94
0.075	107.23	109.23	133.27	165.92
0.100	63.82	66.74	82.85	106.92
0.125	40.27	41.79	52.33	67.29
0.150	27.45	28.87	34.58	43.80
0.175	19.59	20.41	24.34	29.97
0.200	14.92	15.76	18.21	21.94
0.225	12.02	12.33	14.23	16.81
0.250	9.89	10.18	11.69	13.47
	Sensitivity Analysis of Hyper-parameters (Scale)			
	$\kappa_0^2 = 45$, $\kappa_1^2 = 16$	$\kappa_0^2 = 35$, $\kappa_1^2 = 12$	$\kappa_0^2 = 25$, $\kappa_1^2 = 8$	$\kappa_0^2 = 20$, $\kappa_1^2 = 6$
0.000	304.08	291.31	291.04	288.04
0.025	269.99	262.75	258.97	276.15
0.050	182.51	182.96	175.02	200.66
0.075	107.36	107.82	107.31	127.27
0.100	64.93	64.66	63.50	75.37
0.125	40.43	41.43	40.29	47.36
0.150	27.26	28.23	26.85	31.33
0.175	19.85	20.33	19.69	22.27
0.200	14.81	15.31	15.12	16.75
0.225	12.07	12.13	11.83	13.24
0.250	9.97	10.05	9.12	10.78
	Sensitivity Analysis of Hyper-parameters (Variance)			
	$\alpha_0 = 0.1$, $\beta_0 = 0.1$	$\alpha_0 = 0.68$, $\beta_0 = 0.3$	$\alpha_0 = 1.5$, $\beta_0 = 0.9$	$\alpha_0 = 2.5$, $\beta_0 = 2.0$
0.000	307.37	299.96	300.68	297.49
0.025	274.79	259.68	251.64	258.74
0.050	179.01	172.66	167.66	169.51
0.075	106.93	103.49	101.37	99.10
0.100	64.04	61.43	59.70	60.60
0.125	40.68	39.02	38.39	38.00
0.150	27.46	26.27	25.65	26.05
0.175	19.65	19.44	19.00	19.01
0.200	15.02	14.75	14.55	14.59
0.225	11.97	11.65	11.67	11.65
0.250	9.84	9.68	9.57	9.59

Table 3.5: Bayesian ARL_1 values for shifts in variance at $ARL_0 = 300$

δ_V	Sensitivity Analysis of Hyper-parameters (Location)			
	$b_0 = 15$, $b_1 = 1.5$	$b_0 = 20$, $b_1 = 2.5$	$b_0 = 25$, $b_1 = 3.5$	$b_0 = 30$, $b_1 = 4.5$
1.00	313.27	302.41	302.96	297.45
1.20	28.67	28.79	33.31	53.14
1.40	11.00	11.16	12.03	17.31
1.60	6.88	7.08	7.46	9.58
1.80	5.19	5.33	5.49	6.74
2.00	4.25	4.32	4.52	5.29
2.20	3.67	3.75	3.90	4.43
2.40	3.28	3.34	3.47	3.87
2.60	3.03	3.08	3.18	3.52
2.80	2.82	2.89	2.94	3.23
3.00	2.70	2.71	2.81	3.02
	Sensitivity Analysis of Hyper-parameters (Scale)			
	$\kappa_0^2 = 45$, $\kappa_1^2 = 16$	$\kappa_0^2 = 35$, $\kappa_1^2 = 12$	$\kappa_0^2 = 25$, $\kappa_1^2 = 8$	$\kappa_0^2 = 20$, $\kappa_1^2 = 6$
1.00	306.87	300.40	309.71	297.31
1.20	28.66	28.76	29.27	28.85
1.40	11.06	11.00	11.17	11.09
1.60	6.91	6.70	7.03	7.04
1.80	5.14	5.19	5.28	5.27
2.00	4.29	4.29	4.29	4.34
2.20	3.65	3.71	3.70	3.76
2.40	3.01	3.32	3.35	3.38
2.60	3.04	3.06	3.07	3.11
2.80	2.83	2.84	2.87	2.90
3.00	2.67	2.70	2.72	2.75
	Sensitivity Analysis of Hyper-parameters (Variance)			
	$\alpha_0 = 0.1$, $\beta_0 = 0.1$	$\alpha_0 = 0.68$, $\beta_0 = 0.3$	$\alpha_0 = 1.5$, $\beta_0 = 0.9$	$\alpha_0 = 2.5$, $\beta_0 = 2.0$
1.00	304.85	297.20	298.92	290.02
1.20	28.78	27.39	26.91	25.99
1.40	11.14	9.93	9.32	9.19
1.60	6.93	6.03	5.55	5.59
1.80	5.19	4.51	4.18	4.12
2.00	4.24	3.72	3.46	3.40
2.20	3.66	3.26	3.03	3.00
2.40	3.30	2.94	2.75	2.73
2.60	3.04	2.72	2.56	2.57
2.80	2.83	2.56	2.45	2.43
3.00	2.68	2.45	2.35	2.35

Table 3.6: ARL_1 Comparisons of Classical and Bayesian EWMA at $ARL_0 = 300$. (Note that the values after ± are not standard deviations but the boundaries of the 95% confidence interval.)

	Shifts in Y-Intercept	
δ_I	Non-Bayesian EWMA	Bayesian EWMA
0.0	293.11 ±5.93	307.01 ±5.92
0.2	74.98 ±2.81	43.71 ±0.75
0.4	18.47 ±0.71	13.61 ±0.18
0.6	8.67 ±0.22	7.24 ±0.07
0.8	5.53 ±0.09	4.83 ±0.04
1.0	4.14 ±0.05	3.75 ±0.03
1.2	3.36 ±0.03	3.10 ±0.02
1.4	2.88 ±0.02	2.70 ±0.01
1.6	2.57 ±0.01	2.42 ±0.01
1.8	2.35 ±0.01	2.26 ±0.01
2.0	2.20 ±0.01	2.13 ±0.01
	Shifts in Slope	
δ_S	Non-Bayesian EWMA	Bayesian EWMA
0.000	291.27 ±6.00	290.41 ±5.80
0.025	253.50 ±5.56	241.09 ±4.85
0.050	168.75 ±4.69	160.93 ±3.12
0.075	101.82 ±3.51	97.47 ±1.86
0.100	60.15 ±2.46	59.51 ±1.07
0.125	38.38 ±1.70	37.89 ±0.65
0.150	26.40 ±1.09	26.23 ±0.41
0.175	19.22 ±0.78	19.10 ±0.27
0.200	14.54 ±0.53	14.70 ±0.19
0.225	11.72 ±0.37	11.70 ±0.14
0.250	9.65 ±0.28	9.70 ±0.11
	Shifts in Variance	
δ_E	Non-Bayesian EWMA	Bayesian EWMA
1.0	295.25 ±6.00	292.63 ±6.00
1.2	40.34 ±0.80	39.02 ±0.68
1.4	13.03 ±0.24	12.46 ±0.17
1.6	6.98 ±0.11	7.12 ±0.08
1.8	4.77 ±0.06	5.07 ±0.05
2.0	3.69 ±0.04	4.12 ±0.04
2.2	3.19 ±0.03	3.55 ±0.03
2.4	2.82 ±0.02	3.16 ±0.02
2.6	2.61 ±0.02	2.90 ±0.02
2.8	2.46 ±0.02	2.72 ±0.04
3.0	2.35 ±0.01	2.58 ±0.02

Table 3.7: ARL_1 values for T^2 EWMA Regression. (Note that the values after ± are not standard deviations but the boundaries of the 95% confidence interval.)

Shifts					
δ_I	Intercept	δ_S	Slope	δ_V	Variance
0.0	307.22± 5.90	0.000	310.14± 6.04	1.0	307.53± 5.82
0.2	53.20± 1.24	0.025	206.83± 4.92	1.2	57.35± 1.43
0.4	11.72± 0.36	0.050	128.66± 2.91	1.4	22.42± 0.63
0.6	3.99± 0.24	0.075	74.00± 1.76	1.6	11.49± 0.39
0.8	1.44± 0.22	0.100	42.35± 1.03	1.8	6.27± 0.31
1.0	0.27± 0.21	0.125	26.10± 0.66	2.0	4.05± 0.27
1.2	-0.47± 0.20	0.150	17.44± 0.48	2.2	2.31± 0.24
1.4	-0.97± 0.19	0.175	11.89± 0.36	2.4	1.23± 0.22
1.6	-1.21± 0.19	0.200	8.53± 0.31	2.6	0.72± 0.21
1.8	-1.29± 0.19	0.225	6.36± 0.27	2.8	0.20± 0.21
2.0	-1.94± 0.19	0.250	4.61± 0.25	3.0	-0.40± 0.20
15.0	-2.22± 0.19	2.000	-2.35± 0.19	15.0	-2.07± 0.19

Table 3.8: ARL_1 values for T^2 CUSUM Regression, K=0.5. (Note that the values after ± are not standard deviations but the boundaries of the 95% confidence interval.)

Shifts					
δ_I	Intercept	δ_S	Slope	δ_V	Variance
0.0	312.66± 5.95	0.000	313.23± 5.84	1.0	305.41± 5.98
0.2	39.95± 0.90	0.025	194.71± 4.45	1.2	65.22± 1.50
0.4	9.78± 0.30	0.050	105.12± 2.37	1.4	27.55± 0.71
0.6	4.17± 0.22	0.075	56.05± 1.26	1.6	15.46± 0.44
0.8	2.24± 0.22	0.100	32.45± 0.73	1.8	9.49± 0.33
1.0	0.90± 0.20	0.125	20.94± 0.48	2.0	6.57± 0.28
1.2	0.31± 0.19	0.150	13.80± 0.37	2.2	4.10± 0.24
1.4	-0.11± 0.19	0.175	10.45± 0.31	2.4	3.27± 0.23
1.6	-0.32± 0.18	0.200	8.03± 0.27	2.6	2.19± 0.22
1.8	-0.76± 0.18	0.225	6.19± 0.25	2.8	1.64± 0.20
2.0	-0.79± 0.18	0.250	5.00± 0.24	3.0	1.05± 0.20
15.0	-2.03± 0.17	2.000	-1.83± 0.17	15.0	-1.80± 0.17

Table 3.9: ARL_1 values for T^2 CUSUM Regression, K=1. (Note that the values after ± are not standard deviations but the boundaries of the 95% confidence interval.)

Shifts					
δ_I	Intercept	δ_S	Slope	δ_V	Variance
0.0	302.99 ± 5.87	0.000	304.58 ± 5.84	1.0	306.97± 5.92
0.2	76.98± 1.85	0.025	224.47 ± 5.24	1.2	45.06 ± 1.16
0.4	15.58 ± 0.47	0.050	163.75 ± 3.79	1.4	15.09 ± 0.48
0.6	3.84 ± 0.26	0.075	101.51± 2.37	1.6	6.89 ± 0.32
0.8	1.01 ± 0.28	0.100	63.41 ± 1.54	1.8	3.17 ± 0.27
1.0	-0.33 ± 0.21	0.125	39.80 ± 1.00	2.0	1.62 ± 0.24
1.2	-1.14 ± 0.21	0.150	24.82 ± 0.68	2.2	0.34 ± 0.23
1.4	-1.48 ± 0.21	0.175	16.60± 0.50	2.4	-0.22 ± 0.22
1.6	-1.82 ± 0.20	0.200	11.17 ± 0.40	2.6	-0.69 ± 0.21
1.8	-1.67 ± 0.20	0.225	7.44 ± 0.32	2.8	-1.04 ± 0.21
2.0	-2.10 ± 0.20	0.250	5.16 ± 0.29	3.0	-1.38 ± 0.21
15.0	-2.46 ± 0.20	2.000	-2.55 ± 0.19	15.0	-2.59 ± 0.19

3.4 Effect of Varying Values of Smoothing Constants Illustrated with Plots

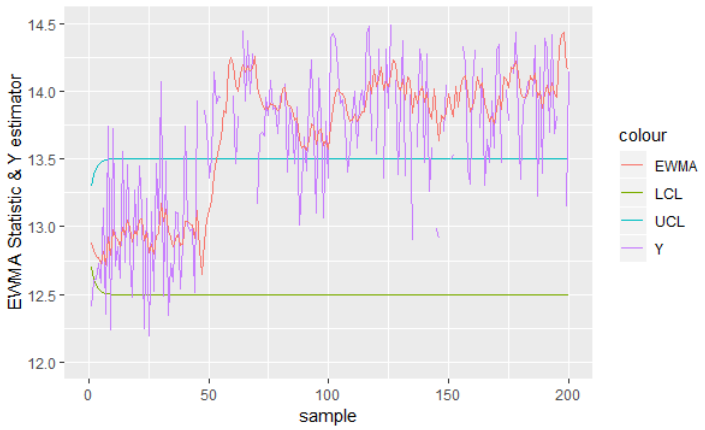

Figure 3.1: Example of a control chart that monitors the Y-intercept estimator which shifted at the 51^{st} sample.

The following plots illustrate the effect of varying the smoothing constant λ on the ability of the EWMA chart to detect different magnitudes of shifts σ. Each data point represent the average of 10,000 data points. Hence, the plots do not show as much variance as in Figure 3.1 which is an example of one single run. In summary, the plots show that larger values of λ is better at detecting larger shifts and vice versa. The number on the left in each chart is the j^{th} sample at which the chart signals when monitoring only the current sample while the number on the rights is the accumulated chart statistic. All shifts begin at the 51^{st} sample.

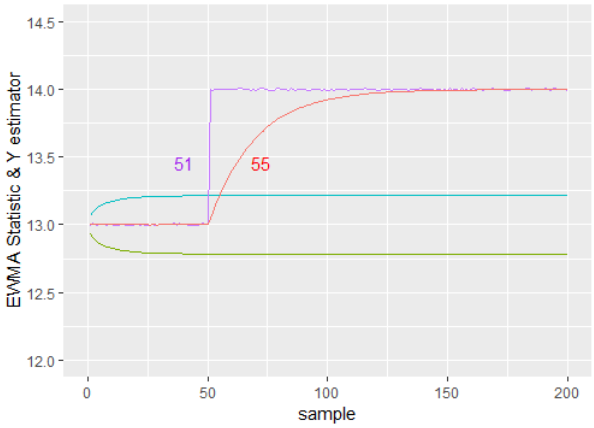

Figure 3.2: $\lambda = 0.05$, Shift=1

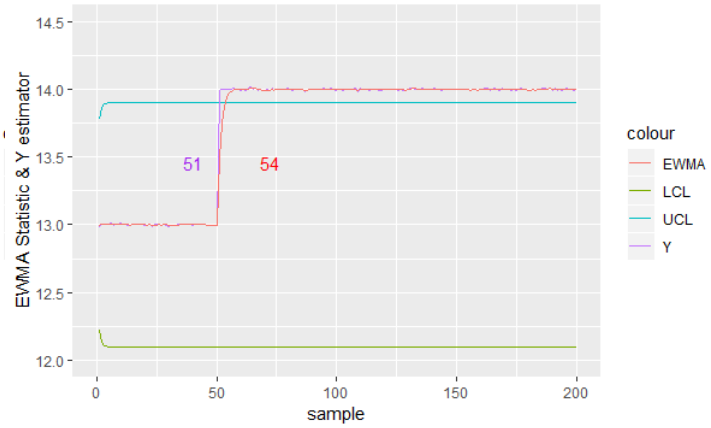

Figure 3.3: $\lambda = 0.5$, Shift=1

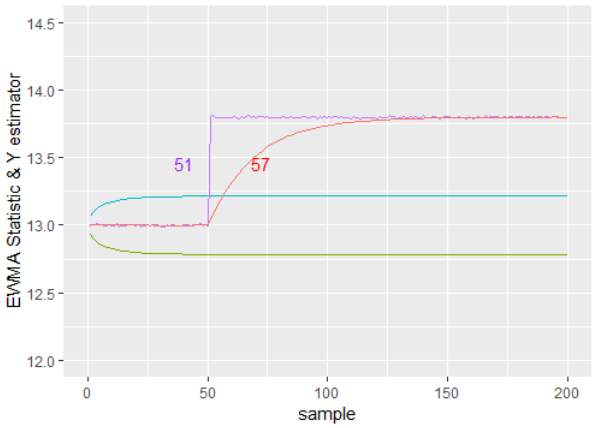

Figure 3.4: $\lambda = 0.05$, Shift=0.8

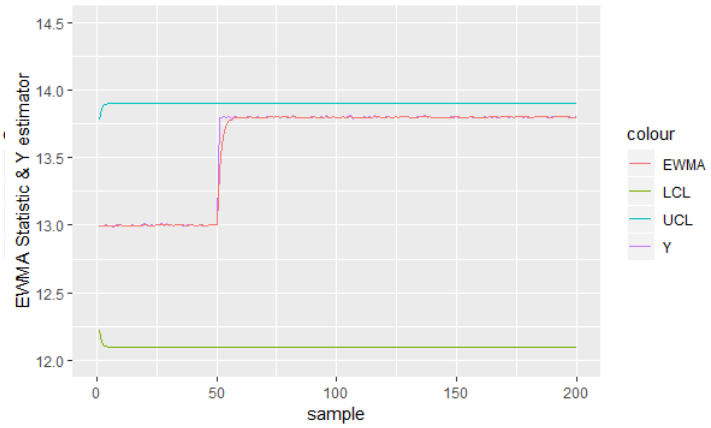

Figure 3.5: $\lambda = 0.5$, Shift=0.8

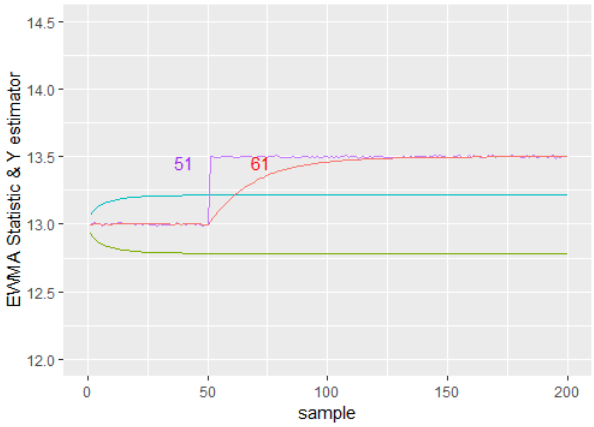

Figure 3.6: $\lambda = 0.05$, Shift=0.5

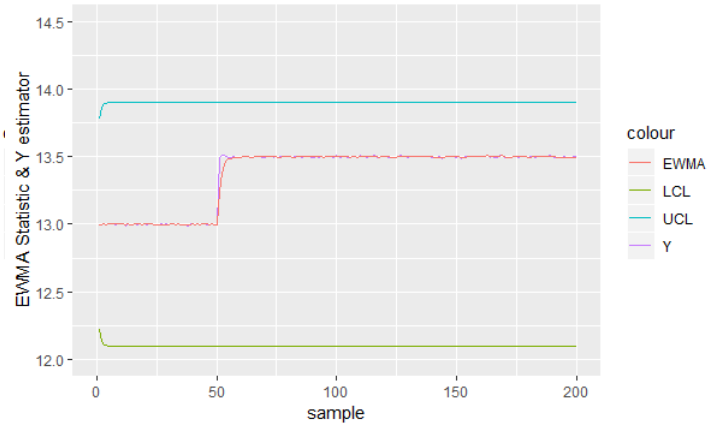

Figure 3.7: $\lambda = 0.5$, Shift=0.5

4

Conclusion and Discussion

Shifts\Type	Non-Bayesian EWMA	Bayesian EWMA	T^2 EWMA	T^2 CUSUM (K=0.5)	T^2 CUSUM (K=1.0)
Intercept	4.59	4.09	0.66	0.74	0.76
Slope	0.50	0.49	0.33	0.28	0.44
Variance	35.85	36.60	37.09	46.83	28.40

Table 4.1: EQL values for all techniques described.

This paper provides an overview of techniques used to construct a control chart that monitors the quality of a process modelled as a linear regression. The parameters of the regression are derived using the ordinary least squares method, as well as using Bayesian inference. Under the EWMA chart, the Bayesian approach performs better for a shift in intercept and slope, but not for variance. The smoothing constant λ of EWMA controls the sensitivity of the chart. The standard value is set to $\lambda = 0.2$ but greater values are better in detecting large shifts and vice versa. The T^2 control chart also performs better in general compared to using 3 separate control charts to monitor the parameters individually. However, the disadvantage of the T^2 chart is that the cause of instability of the process when it occurs cannot be determined since the information from all estimators is combined into a single value. The T^2 CUSUM chart for $K = 0.5$ performs better than the T^2 EWMA chart when the shifts are small (as shown in the slope shifts). A greater value of $K = 1$ is better at detecting variance shifts (measures in multiples of the original variance). Hence, the parameter K in the CUSUM chart serves the same purpose as λ in the EWMA chart, i.e. they determine the sensitivity of the charts. A summary of results is presented in Table 4.1, lower EQL values indicate better performance. Since there is no analytic expression for ARL in this paper, the EQL values are estimated using the trapezoid rule with the AUC function of the $MESS$ library in R. A shift in slope is always the earliest to be detected while a shift in process variance is the hardest to detect for all charts. In conclusion, there is no one chart that performs strictly better than the others. Hence, the preferred choice of control chart should depend on the type of process to be monitored and the required sensitivity.

The steps in deriving the Bayesian estimators of the parameters are followed in close reference to the paper by Abbas et. al. [2]. However, there are some notations that are not explained fully and hence require some assumptions. For example, in Section 2.5 Equation 18, it is unclear whether "σ^2" refers to the true value or the posterior

parameter. The latter is assumed to be true. The model used in the referred paper was also not specified even though is was hinted that they used the model by Kim et. al. [13] in the first paragraph of Section 2, so the model used in this paper is also taken to be $Y = 13 + 2X + \epsilon$. Consequently, the results of this paper and the (relevant part of the) referred paper are not entirely similar, but the general trend is still observed. There are two parameters in each chart that affect their sensitivity and performance. In the EWMA chart, the parameters are denoted λ and L while in the CUSUM chart, the parameters are K and h. When varying the values of λ and K, the value of ARL_0 is affected. Hence, the other parameters L and h need to be adjusted accordingly to maintain a fix value of ARL_0. This is the reason why in Table 3.2-3.7, even though there is a jump in every chart at the 51^{st} sample, some do not exceed the limits because the width of the chart has been adjusted to maintain an ARL_0 of approximately 300.

There are some aspects for future improvement in this paper. Firstly, the derivation of control limits given in Section 2.5 could be made more efficient and precise. The method used in this paper is based on estimation and manual adjustment which might not be feasible when there are many parameters. Secondly, the comparison is based solely on computer simulations with the R software due to a lack of real data. Before a method is chosen to be used on a real-world process, they should first be tested on actual data sets. Next, it is assumed that the real or ideal values of the model parameters are known, further study could be done to modify the current techniques when the actual parameters are unknown. The scope of this paper could be expanded by exploring a non-linear relationship between the variables of the process such as polynomial functions, as well as extending the current techniques to multivariate processes.

Bibliography

[1] Abbas, N., Riaz, M., and Does, R. J. M. M. (2014). An EWMA-Type Control Chart for Monitoring the Process Mean Using Auxiliary Information . *Communications in Statistics - Theory and Methods*, 43(16):3485–3498.

[2] Abbas, T., Ahmad, S., Riaz, M., and Qian, Z. (2016). On Monitoring of Linear Profiles Using Bayesian Methods. *Computers & Industrial Engineering*, 94:245–268.

[3] B., C. R. (1988). Multivariate Generalizations of Cumulative Sum Quality-Control Schemes. *Technometrics*, 30(3):291–303.

[4] Baradaran, V. and Dashtbani, H. (2019). A decision support system for monitoring traffic by statistical control charts- scientific figure on researchgate. [Online; accessed 8 Apr, 2019].

[5] Cochran, W. G. (1977). *Sampling Techniques (3rd ed.)*. John Wiley & Sons, New York, NY.

[6] Coelho, M., Chakraborti, S., and Graham, M. (2015). A Comparison of Phase I Control Charts. *South African Journal of Industrial Engineering*, 26:178 – 190.

[7] Crowder, S. V. and Hamilton, M. D. (1992). An EWMA for Monitoring a Process Standard Deviation. *Journal of Quality Technology*, 24:12–21.

[8] Deming, W. E. walter a. shewhart, 1891-1967. *Revue de l'Institut International de Statistique / Review of the International Statistical Institute*, (3):372–375.

[9] DiPaola, P. P. (1945). Use of Correlation in Quality Control. *Industrial Quality Control*, 2:10–14.

[10] Grey, K. (2019). XmR Control Limits | Why Moving Range, not Standard Deviation. [Online; accessed 9 Apr, 2019].

[11] Kalgonda, A.A and Koshti, V.V and Ashokan, K.V and Kolhapur and Mahankal, K. and Maharashtra, K. (2011). Exponentially weighted moving average control chart. *Asian Journal of Management Research*.

[12] Kang, L. and Albin, S. L. (2000). On-Line Monitoring When the Process Yields a Linear Profile. *Journal of Quality Technology*, 32:418–426.

[13] Kim, K., A., M., and Woodall, W. H. (2003). On the Monitoring of Linear Profiles. *Journal of Quality Technology*, 35:317–328.

[14] Lawless, J. (1980). *Statistical Models and Methods for Lifetime Data*. John Wiley & Sons, New York, NY.

[15] Mandel, B. (1969). The Regression Control Chart. *Journal of Quality Technology*, 1:1–9.

[16] Montgomery, D. (2013). *Introduction to Statistical Quality Control*. John Wiley Sons, Inc., United States.

[17] Noorossana, R., Eyvazian, M., and Vaghefi, A. (2010). Phase II Monitoring of Multivariate Simple Linear Profiles. *Computers & Industrial Engineering*, 58(4):563–570.

[18] Page, E. S. (1954). Continuous Inspection Schemes. *Biometrika*, 41:100–115.

[19] Roberts, S. (1959). Control Chart Tests Based on Geometric Moving Averages. *Technometrics*, 3:349–364.

[20] Z., L., Zou, C., Gong, Z., and Wang, Z. (2014). The Computation of Average Run Length and Average Time to Signal: An Overview . *Journal of Statistical Computation and Simulation*, 84:8:1779–1802.

A

Appendix A

A.1

From Equation (2.9):

$$\begin{aligned} P(B_0, B_1|Y) \propto \exp(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_{ij} - B_0 - B_1 X_i')^2) \\ \times \ \exp(-\frac{1}{2\kappa_0^2}(B_0 - b_0)^2) \\ \times \ \exp(-\frac{1}{2\kappa_1^2}(B_1 - b_1)^2) \end{aligned}$$

Expanding square terms and ignoring fixed parameters,

$$\begin{aligned} P(B_0, B_1|Y) \propto \exp(-\frac{1}{2\sigma^2}(nB_0^2 - 2B_0\sum_{i=1}^{n} y_{ij} + B_1^2\sum_{i=1}^{n} X_i'^2 - 2B_1\sum_{i=1}^{n} X_i' y_{ij})) \\ \times \ \exp(-\frac{1}{2\kappa_0^2}(B_0^2 - B_0 b_0 + b_0^2)) \times \ \exp(-\frac{1}{2\kappa_1^2}(B_1^2 - 2B_1 b_1 + b_1^2)) \end{aligned}$$

Gathering terms,

$$\begin{aligned} P(B_0, B_1|Y) \propto \exp((-\frac{n}{2\sigma^2} - \frac{1}{2\kappa_0^2})B_0^2 + (\frac{\sum_{i=1}^{n} y_{ij}}{\sigma^2} + \frac{b_0}{\kappa_0^2})B_0 \\ + (\frac{\sum_{i=1}^{n} x_i y_{ij}}{\sigma^2} + \frac{b_1}{\kappa_1})B_1 + (-\frac{\sum_{i=1}^{m} x_i^2}{2\sigma^2} - \frac{1}{2\kappa_1^2})B_1^2) \end{aligned}$$

B_0 and B_1 are independent and so can be separated out,

$$\begin{aligned} P(B_0|Y) \propto \ \exp(-\frac{1}{2}\frac{n\kappa_0^2 + \sigma^2}{\sigma\kappa_0^2}B_0^2 + \frac{n\bar{y}_j\kappa_0^2 + b_0\sigma^2}{\sigma^2\kappa_0^2}B_0) \\ \propto \ \exp(-\frac{1}{2}\frac{n\kappa_0^2 + \sigma^2}{\sigma\kappa_0^2}[B_0^2 + \frac{n\bar{y}_j\kappa_0^2 + b_0\sigma^2}{n\kappa_0^2 + \sigma^2}]^2) \\ \propto \ \exp(-\frac{1}{2}\frac{1}{\sigma_{0nj}^2}(B_0 - b_{0nj})^2) \end{aligned}$$

The same process is repeated for the other parameters.

A.2

Elements of Σ as mentioned in Section 2.4:

$$Cov(\hat{\beta_{0k}}, \hat{\beta_{0k}}) = \sigma^2(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}})$$

$$Cov(\hat{\beta_{1k}}, \hat{\beta_{1k}}) = \frac{\sigma}{S_{xx}}$$

$$Cov(\hat{\beta_{0k}}, \hat{\beta_{1k}}) = -\sigma\frac{\bar{x}}{S_{xx}}$$

A.3

This is the R code that was written for generating the results reported in Section 3.1. ρ (rho) represents the correlation value which is varied according to Table 3.1. The target value values μ_y (mu_y) and μ_x (mu_x) are set to 0 while δ (delta) = 0.5, $L = 2.824$, σ_y (s_y) = 1, σ_x (s_x) = 1 and they refer to the variance of Y and X respectively. The code is written as a function such that these values could be changed without changing the code.

```
bivar_function <-function(rho,delta,L,N,mu_y,mu_x,s_y,s_x,lambda){

    b_YX<-rho*(s_y/s_x)
    sigma <- matrix(c(s_y^2, s_y*s_x*rho, s_y*s_x*rho, s_x^2),
                    2) # Covariance matrix
    mu1 <- c(mu_y,mu_x)    # In control mean
    mu2 <- c(mu_y + delta*s_y, mu_x)
    ARL<-vector()
    ARL1<-vector()

    for (j in 1:10000){

        samples<-matrix(ncol=2, nrow=30)
         for (i in 1:10){    #simulate IC samples
            obs<-mvrnorm(N, mu = mu1, Sigma = sigma )
            samples[i,1]<-mean(obs[,1])
            samples[i,2]<-mean(obs[,2])   }
         for (i in 11:30){    #simulate OC samples
            obs<-mvrnorm(N, mu = mu2, Sigma = sigma )
            samples[i,1]<-mean(obs[,1])
            samples[i,2]<-mean(obs[,2])   }


    M<-samples[,1] + b_YX*(mu_x-samples[,2])   #regression estimate
    LCL<-vector()
    UCL<-vector()
```

```
Z<- vector()
Z_old<-mu_y
var_M<-(s_y^2-(b_YX^2)*(s_x^2))/N
Z1<-vector()
LCL1<-vector()
UCL1<-vector()
Z1_old<-mu_y
arl<-30
arl1<-30
for (i in 1:30){
#with auxiliary

    Z[i] <- lambda*M[i] + (1-lambda)*Z_old
    UCL[i]<-mu_y+ L*sqrt(var_M)
          *sqrt((lambda/(2-lambda))*(1-(1-lambda)^(2*i)))
    LCL[i]<-mu_y- L*sqrt(var_M)
          *sqrt((lambda/(2-lambda))*(1-(1-lambda)^(2*i)))
    Z_old<-Z[i]
    if (Z[i]>UCL[i] || Z[i]<LCL[i] ) {
        arl<-min(arl,i)
        }

 #without auxiliary
    Z1[i] <- lambda*samples[i,1] + (1-lambda)*Z1_old
    UCL1[i]<-mu_y+ L*s_y
              *sqrt((lambda/((2-lambda)*N))*(1-(1-lambda)^(2*i)))
    LCL1[i]<-mu_y- L*s_y
             *sqrt((lambda/((2-lambda)*N))\\*(1-(1-lambda)^(2*i)))
    Z1_old<-Z1[i]
    if (Z1[i]>UCL1[i] || Z1[i]<LCL1[i] ) {
        arl1<-min(arl1,i)
        }

    }
ARL[j]<-arl
ARL1[j]<-arl1}

results<-cbind(mean(ARL)-10, mean(ARL1)-10)
return(results)
 }
```

A.4

Below are two sets of codes mentioned in Section 3.2. The first one is the non-Bayesian approach where the estimators are estimated by OLS method while the

second is the Bayesian approach. The magnitude of shifts are presented in Table 3.2.

For the non-Bayesian approach, the code is initialized as follows, the values of L_I, L_S, L_V that are shown are set such that $ARL_0 \approx 300$ for λ (lambda) $= 0.2$:

```
X1<- c(-3,-1,1,3)
k<- 10000
n<- length(X1)
L_I<- 3.15
L_S<- 3.15
L_V<- 1.45
lambda<-0.2

non_Bayesian_EWMA_function<-function(X1,lambda,intercept_shift,
slope_shift,error_shift,L_I,L_S,L_V,k,n){

  ARL_I<-vector()
  S_xx<-sum((X1-mean(X1))^2)

  for (j in 1:k){     #k is the number of runs
        Y2<-vector()
        Y3<-vector()
        Y4<-vector()

        for (i in 1:100){
            y_j<-13+2*X1+rnorm(n,0,1)
            Y2[i]<-mean(y_j)
            S_xy<-sum((X1-mean(X1))*y_j)
            Y3[i]<-S_xy/S_xx
            y1_j<-y_j- (Y2[i]+Y3[i]*X1)
            Y4[i]<-sum((y1_j^2))/(n-2)
            }

        for(i in 101:1000){  #shift happens at the 11th sample
            y_j<-(13+intercept_shift)+(2+slope_shift)
                 *X1+rnorm(n,0,error_shift)
            Y2[i]<-mean(y_j)
            S_xy<-sum((X1-mean(X1))*y_j)
            Y3[i]<-S_xy/S_xx
            y1_j<-y_j- (Y2[i]+Y3[i]*X1)
            Y4[i]<-sum((y1_j^2))/(n-2)
            }


    z_I_old<-13
    z_I<-vector()
```

```
  z_S_old<-2
  z_S<-vector()
  z_V_old<-0
  z_V<-vector()
  LCL_I<-vector()
  UCL_I<-vector()
  LCL_S<-vector()
  UCL_S<-vector()
  UCL_V<-vector()
  Var<-(2/(n-2))+(2/((n-2)^2))+(4/(3*(n-2)^3))-(16/(15*(n-2)^5))
  ARL<-1000

  for (i in 1:1000){
    z_I[i]<-theta*Y2[i]+(1-lambda)*z_I_old
    z_I_old<-z_I[i]
    LCL_I[i]<-13-L_I*sqrt(lambda/((2-lambda)*4)*(1-(1-lambda)^(2*i)))
    UCL_I[i]<-13+L_I*sqrt(lambda/((2-lambda)*4)*(1-(1-lambda)^(2*i)))

    z_S[i]<-theta*Y3[i]+(1-theta)*z_S_old
    z_S_old<-z_S[i]
    LCL_S[i]<-2-L_S*sqrt(lambda/((2-lambda)*S_xx)*(1-(1-lambda)^(2*i)))
    UCL_S[i]<-2+L_S*sqrt(lambda/((2-lambda)*S_xx)*(1-(1-lambda)^(2*i)))

    z_V[i]<-max(lambda*log(Y4[i])+(1-lambda)*z_V_old,log(1))
    z_V_old<-z_E[i]
    UCL_V[i]<-L_V*sqrt((theta/(2-lambda))*(1-(1-lambda)^(2*i))*Var)
    if (z_I[i]>UCL_I[i] ||
      z_I[i]<LCL_I[i] ||
      z_S[i]>UCL_S[i] ||
      z_S[i]<LCL_S[i] ||
      z_V[i]>UCL_V[i]) {
          ARL<-min(ARL,i)
    }
  }
  ARL_I[j]<-ARL  }
mean(ARL_I)-100  }
```

As for the Bayesian approach, the hyper-parameters and shifts are varied according to Table 3.3-3.5, an example of the initial values are as follows, L_I, L_S, L_V are set such that $ARL_0 \approx 300$ for λ (lambda) = 0.2:

```
X1<-c(-3,-1,1,3)
n<-length(X1)
lambda<-0.2
K_0_squared<-45
K_1_squared<-16
b0<-20
b1<-2.5
alpha_0<-0.1
beta_0<-0.1
k<-10000
L_I<- 3.30
L_S<- 3.23
L_V<- 2.64

Bayesian_function <-function(X1,lambda,intercept_shift,
            slope_shift,error_shift,L_I,L_S,L_V,
            k,n,K_0_squared,K_1_squared,alpha_0,beta_0){

  ARL_b_I<-vector()
  S_xx<-sum(X1^2)

  for (j in 1:k){
    Y4<-vector()
    b0<-vector()
    b1<-vector()

    for (i in 1:100){
      y_j<-13+2*X1+rnorm(n,0,1)
      Y<-mean(y_j)
      sigma0_2<-sum((y_j-(13+2*X1))^2)/(n-2)
      sigma_2<-(beta_0+((n-2)*sigma0_2)/2)/(alpha_0+(n/2)-1)
      S_xy<-sum(X1*y_j)
      b0[i]<-((4*Y*K_0_squared)+15*sigma_2)
            /(4*K_0_squared + sigma_2)
      b1[i]<-((S_xy*K_1_squared)+1.5*sigma_2)
            /(S_xx*K_1_squared + sigma_2)
      Y4[i]<- sigma_2}


    for(i in 101:1000){  #shift happens at the 101th sample
      y_j<-(13+intercept_shift)+(2+slope_shift)*X1
            +rnorm(n,0,error_shift)
```

```
    Y<-mean(y_j)
    sigma0_2<-sum((y_j-(13+2*X1))^2)/(n-2)
    sigma_2<-(beta_0+((n-2)*sigma0_2)/2)/(alpha_0+(n/2)-1)
    S_xy<-sum(X1*y_j)
    b0[i]<-((4*Y*K_0_squared)+15*sigma_2)
          /(4*K_0_squared + sigma_2)
    b1[i]<-((S_xy*K_1_squared)+1.5*sigma_2)
          /(S_xx*K_1_squared + sigma_2)
    Y4[i]<-sigma_2
  }

  z_old<-13
  z_S_old<-2
  z_E_old<-0
  z_I<-vector()
  z_S<-vector()
  z_V<-vector()
  LCL_I<-vector()
  UCL_I<-vector()
  LCL_S<-vector()
  UCL_S<-vector()
  UCL_V<-vector()       #only upper limit to detect
increase in variance

  ARL<-1000
  Q1<-2*alpha_0 + (n-2)
  Q2<-2*alpha_0^2 + (n-2)^2 + 4*alpha_0*(n-2)
  Q3<-8*(alpha_0^3) + 12*(alpha_0^2)*(n-2)
      + 6*alpha_0*(n-2)^2 + (n-2)^3
  Q4<-32*(alpha_0^5)
      + 80*(alpha_0^4)*(n-2) + 80*(alpha_0^3)*(n-2)^2
      + 40*(alpha_0^2)*(n-2)^3 + 10*alpha_0*(n-2)^4 + (n-2)^5

  Var<-(2/Q1)+(2/Q2)+(4/(3*Q3))-(16/(15*Q4))

  for (i in 1:1000){
    z_I[i]<-lambda*b0[i] + (1-lambda)*z_old
    UCL_I[i]<-13+L_I*sqrt(((lambda*K_0_squared)/((2-lambda)
              *(4*K_0_squared+1)))*(1-(1-lambda)^(2*i)))
    LCL_I[i]<-13-L_I*sqrt(((lambda*K_0_squared)/((2-lambda)
              *(4*K_0_squared+1)))*(1-(1-lambda)^(2*i)))
    z_old<-z_I[i]

    z_S[i]<-lambda*b1[i] + (1-lambda)*z_S_old
    UCL_S[i]<-2+ L_S*sqrt(((lambda*K_1_squared)/((2-lambda)
              *(S_xx*K_1_squared+1)))*(1-(1-lambda)^(2*i)))
```

```
      LCL_S[i]<-2- L_S*sqrt(((lambda*K_1_squared)/((2-lambda)
                  *(S_xx*K_1_squared+1)))*(1-(1-lambda)^(2*i)))
      z_S_old<-z_S[i]

      z_V[i]<-max(lambda*log(Y4[i])+(1-lambda)*z_E_old,log(1))
      z_V_old<-z_E[i]
      UCL_V[i]<-L_E*sqrt((lambda/(2-lambda))*(1-(1-lambda)^(2*i))*Var)
      if (z_I[i]>UCL_I[i] ||
        z_I[i]<LCL_I[i] ||
        z_S[i]>UCL_S[i] ||
        z_S[i]<LCL_S[i] ||
        z_V[i]>UCL_E[i]) {
            ARL<-i
            break
        }
    }
    ARL_b_I[j]<-ARL   }
mean(ARL_b_I)-100   }
```