A STUDY OF EXPONENTIALLY WEIGHTED MOVING AVERAGE (EWMA) METHODOLOGIES

by

LEE TOH TONG

Thesis submitted in fulfilment of the requirements for the degree of Master of Science

Dec 1993

ACKNOWLEDGEMENTS

I am very much indebted to my supervisor, Associate Professor Dr. Quah Soon Hoe for helpful guidance and advice throughout the years.

This research would not have been possible without the facilities provided by Universiti Sains Malaysia, and the School of Mathematical and Computer Sciences, in particular.

Special appreciation has to be tendered to a number of postgraduate students from the School of Mathematical and Computer Sciences for expressing the views from their respective fields, which I have found to be very beneficial and constructive.

I would like to dedicate this thesis to my beloved parents for their support. They have shown the encouragement that has lifted me through some of the discouraging moments that can occur in an activity such as thesis writing.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRAK

SUATU KAJIAN METODOLOGI PURATA BERGERAK BERPEMBERAT EKSPONEN (PBBE)

Tesis ini adalah berkenaan dengan penggunaan carta PBBE, suatu alternatif kepada carta Shewhart dan Carta Hasil Tambah Longgokan (HTL), untuk mengesan perubahan yang berlaku di dalam suatu proses. Kami menulis tesis ini berdasarkan dua objektif:

i) untuk meninjau metodologi carta kawalan PBBE, dan

ii) untuk mencadangkan beberapa kaedah tambahan yang boleh meningkatkan lagi keupayaan carta PBBE.

Di dalam tesis ini, kami memberikan suatu tinjauan secara menyeluruh untuk carta PBBE untuk tujuan mengawal min dan varians suatu proses. Kaedah-kaedah yang boleh meningkatkan keupayaan carta PBBE yang diberi oleh berbagai penulis untuk kedua-dua min dan varians juga akan dibincang. Ketika meninjau semula semua kaedah berlainan yang telah dicadangkan oleh para penulis, kami akan membandingkan setiap satu kaedah yang mempunyai tujuan yang sama dan cuba meringkaskan keserupaan dan perbezaan di antara kaedah-kaedah tersebut. Kebaikan dan keburukan setiap kaedah juga akan dibincang. Kami juga menunjukkan bagaimana carta PBBE dibina dengan menggunakan kaedah yang boleh meningkatkan keupayaan carta PBBE.

Beberapa perbincangan secara am berhubung dengan isu-isu yang berkaitan dengan carta PBBE juga diberikan. Perbincangan berkenaan dengan kebaikan dan keburukan carta PBBE dan perbandingan carta PBBE dengan carta HTL akan membantu kita dalam masalah menentukan sama ada untuk mengguna carta PBBE. Kami juga memberikan suatu perbincangan ringkas berhubung dengan penggunaan PBBE untuk process yang mempunyai min yang tidak tetap, beberapa bidang yang boleh diceburi untuk kajian selanjutnya dan perkembangan masa kini dan prospek carta PBBE.

Di samping meninjau semula beberapa kaedah yang dicadangkan oleh para penulis, kami juga mengemukakan cadangan kami sendiri. Sebagai contoh, melakarkan sisa bersama-sama dengan carta PBBE untuk mengesan perubahan proses min yang besar atau keadaan lain yang abnormal, suatu kaedah yang lebih berkesan untuk mengawasi perubahan proses min yang diiringi oleh kemerosotan varians, dan suatu kaedah untuk menentukan bila dan besarnya perubahan yang telah berlaku kepada suatu proses.

ABSTRACT

This thesis concerns the use of the EWMA chart, an alternative to the Shewhart and Cumulative Sum (CUSUM) charts, for detecting a change in a process. We have written this thesis with two objectives in mind:

i) to review the control charting methodologies of the EWMA control chart, and

ii) to suggest some additional enhancements that can further enhance the EWMA chart.

In this thesis, we provide an overview of the EWMA charts for the monitoring of the process mean and the process variance. Enhancements to the EWMA given by various authors for both the process mean and the process variance are also discussed. While reviewing all the different enhancements proposed by various authors, we will compare each enhancement to the others which serve similar purposes and try to summarize the similarities and differences between these methods. The pros and cons of each of them will also be discussed. We also show how the EWMA chart can be constructed using the appropriate enhancements.

Some general discussions on issues related to the EWMA chart are also given. The discussion on the advantages and weaknesses of the EWMA chart and a comparison of the EWMA chart with the CUSUM chart will enable us to decide whether or not to use the EWMA chart. We also include a brief discussion on the use of the EWMA for a nonstationary process, some possible directions for future research, and current developments and prospects for the EWMA chart.

In addition to a review of some enhancements suggested by various authors, we have also given some suggestions of our own. For example, the plotting of residuals in conjunction with an EWMA chart to detect large shifts in the process mean or any other abnormality, a method which is more effective in detecting a shift in the process mean which causes a reduction in the dispersion, and a method for determining when changes in the process mean start occurring and the magnitudes of such changes.

CHAPTER ONE

THE EWMA FOR MONITORING THE PROCESS MEAN

1.1 Introduction

The Exponentially Weighted Moving Average, EWMA, is already well established in areas like economics, inventory control, forecasting, etc. It was first introduced by Roberts (1959) as a process monitoring and control method alternative to the Shewhart $\bar{X}$ chart and the CUSUM chart. He showed using simulations that the EWMA is superior in detecting small shifts in the mean compared to the traditional Shewhart control chart. This method, however, did not attract the attention from others that it richly deserves until Hunter (1986) raised this issue again by showing that the EWMA can provide a forecast of the next observation. Hence, it can be used in real time dynamic process control.

Basically, an EWMA chart involves the transformation of each sample mean or individual observation into an EWMA statistic before it is plotted on a control chart. The successive values of the EWMA statistic are based on

$$Z_t = \lambda * X_t + (1-\lambda) * Z_{t-1} ,$$
$$0 < \lambda \le 1, \ t = 1,2,\ldots , \qquad (1.1)$$

where X_t = the sample mean or individual observation at time t,

Z_{t-1} = predicted value at time t-1,

Z_t = predicted value at time t, and

λ = smoothing constant.

Z_0 is the starting value. The choice of Z_0 can be very crucial during the start-up of the EWMA control scheme, especially if the λ chosen is very small. For $\lambda = 0.152$, which is optimal in detecting a shift of one σ in the process mean for an in-control ARL of 250 (see Crowder(1989) and Gan (1991B)), Z_0 contributes 85% to the EWMA for Z_1. We will discuss what we mean by an in-control ARL in the next section. Even after the 5th observation, 44% of Z_5 is still due to Z_0. Z_0 could be the target value, the actual mean calculated based on historical data available or even a value called head start which will be discussed later.

X_t can be

i) an individual observed value of a quality characteristic sequentially recorded from a manufacturing process, or

ii) a sample average obtained for a quality characteristic from a designated sampling plan.

Although the X_t's can be independent and identically distributed random variables (IID) from any distribution, normally we assume that the observation is normally distributed with mean μ and variance σ^2, respectively, and can be modelled by

$$X_t = \mu + \varepsilon_t \ , \qquad (1.2)$$

where X_t is the measurement of the process variable at time t,

μ is the mean of the process, and

ε_t is the error term (we assume that the errors are normally and independently distributed with mean, $\mu=0$ and variance, $\sigma^2=1$, frequently written as NID(0,1).)

The choice of λ plays the most important role in EWMA. It determines the rate of response to any process change. However, selecting the value of λ is a tradeoff. A larger λ means that the current observation contributes more to the EWMA. This causes it to respond rapidly when large scale process shifts take place but it is not so efficent in detecting small process changes. A smaller value of λ, on the other hand, can detect small process shifts earlier but is not as good as a large value of λ when large shifts occur. This will be discussed in detail in the next section.

Note also that the sum of the weights is unity, i.e.,

$$\lambda \sum_{t=1}^{m} (1-\lambda)^{m-t} + (1-\lambda)^{m} = 1 \quad , \qquad (1.3)$$

where m is the number of sample means or individual observations.

Besides equation (1.1), the EWMA can also be written in two other forms, namely,

$$\begin{aligned} Z_t &= Z_{t-1} + \lambda * [\ X_t - Z_{t-1}\ ] \\ &= Z_{t-1} + \lambda * \varepsilon_t \ , \qquad (1.4) \end{aligned}$$

where ε_t is the forecast error.

Therefore, the EWMA statistic can also be defined as the previous forecasted value plus a fraction of the forecasting error.

The EWMA is sometimes referred to as a Geometric moving average because Z_t can be written as a weighted average of current and past observations:

$$\begin{aligned} Z_t &= \lambda * X_t + (1-\lambda) * [\ Z_{t-1}\ ] \\ &= \lambda * X_t + (1-\lambda) * [\ \lambda * X_{t-1} + (1-\lambda) * Z_{t-2}\ ] \\ &\ \vdots \\ &= \lambda \sum_{i=0}^{t-1} (1-\lambda)^{i} * X_{t-i} + (1-\lambda)^{t} * Z_0 \qquad (1.5) \end{aligned}$$

It can be noticed that the weightage decreases exponentially as in a geometric series as the observations become more distant from the past. The weightage of the past 3 observations for $\lambda = 0.152$ are as follows:

Table 1.1: Weightages of the EWMA for the Past 3 Observations

Observation	Weightage
X_t	$\lambda = 0.152$
X_{t-1}	$\lambda(1-\lambda) = 0.129$
X_{t-2}	$\lambda(1-\lambda)^2 = 0.109$
X_{t-3}	$\lambda(1-\lambda)^3 = 0.093$

When the X_t is IID with variance σ_X^2, the variance of the EWMA statistic, σ_Z^2 at time t is given by (see Lucas and Saccucci (1990))

$$\sigma_Z^2 = \left\{ \frac{\lambda}{2-\lambda} * \left[1-(1-\lambda)^{2t}\right] \right\} \sigma_X^2 \qquad (1.6)$$

Unless λ is small, the variance quickly converges to its asymptotic value

$$\sigma_Z^2 = \left\{ \frac{\lambda}{2-\lambda} \right\} \sigma_X^2 \qquad (1.7)$$

The objective of this control method is to monitor the mean of a normally distributed process that may experience changes in its mean from the target value. We prefer to use the word *change* instead of shift because we should consider not only abrupt shifts in the process mean but also other cases like gradual drifts.

The control rule for the EWMA is the same as for a Shewhart $\bar{X}$ chart. For a process with an in-control process mean of zero and a standard deviation of one, it is considered to be out-of-control once the EWMA exceeds its control limit, h, i.e., if $|Z_{t+1}| > h$,

where $h = L * \sqrt{\frac{\lambda}{2-\lambda}}$, and (1.8)

L = control limit multiple for the EWMA chart (as defined by Crowder (1987B)).

In this case, the process should be stopped and corrective action should be taken before we are allowed to resume the process again. The choice of h, however, is different from that for the Shewhart $\bar{X}$ chart. This will be discussed in the next section.

Hunter (1986) pointed out that the EWMA can be thought of as a compromise between the Shewhart and Cumulative Sum (CUSUM) control charts. For $\lambda = 1$, EWMA places all its weight on the most recent observation , i.e., it is actually a Shewhart $\bar{X}$ charting method. As the value of λ gets closer to 0, the most recent observation receives a weightage which is not much different from the past few observations as shown in Table 1.1, and the EWMA resembles the CUSUM. Therefore, large values of λ are optimal for detecting large shifts, while small values of λ are optimal for detecting small shifts.

We would like to make it clear that when we refer to a Shewhart chart, it would be to the traditional Shewhart control chart without the supplementary run tests. Incoporating run tests into the Shewhart chart definitely improves the effectiveness of the Shewhart chart but it also reduces its simplicity and makes the interpretation harder.

1.2 Design Strategy

The design of an EWMA involves the choice of two parameters: the smoothing constant, λ, and the control limit, h.

Since the EWMA can be used for both process monitoring and forecasting, and the designs for both purposes differ from each other, one should understand the nature of the process before deciding which strategy to use. Lucas and Saccucci (1990) have shown how to make use of the EWMA for detecting shifts in the mean level more effectively for the white noise process. Montgomery and Mastrangelo (1991) and Box and Kramer (1992) have used the EWMA as a forecasting tool for nonstationary drifts in the process mean.

The design of the EWMA scheme for process monitoring and the detection of shifts depends on its Average Run Length (ARL). The ARL is defined as the average number of sample means required for the EWMA to exceed the control limits for the first time. An ideal control scheme should have an ARL which is as large as possible when the process is in statistical control so that the process engineer does not have to bother with false alarms, and should have an ARL which is as small as possible when the process is out of statistical control so that action can be taken immediately. We will use the terms "in-control ARL" and "out-of-control ARL" to represent these two types of ARL throughout this discussion.

Crowder (1987A) has obtained an integral equation for the ARL of an EWMA chart, which is given below:

$$L(u) = 1 + \frac{1}{\lambda} \int_{-h}^{h} L(x)\ f\left\{ \frac{x-(1-\lambda)u}{\lambda} \right\} dx \qquad (1.9)$$

where L(u) denotes the ARL of an EWMA chart beginning at $Z_0 = u$, $-h<u<h$, and

f(x) is the probability density function of a sample mean, or an individual observation.

The program for computing the ARL using the above method (written in the C programming language) is given in Appendix A (see Crowder (1987B)).

The design of the EWMA for monitoring the mean level of the white noise process which will be discussed here makes use of the ARL suggested by Lucas and Saccucci (1990) and the sensitivity analysis by Robinson and Ho (1978). This involves the following four steps:

i) Choose the smallest acceptable ARL when the process is in control. In other words, we specify the average number of observations before an out-of-control signal is given when the process is actually in control. This is equivalent to selecting an acceptable Type I error in hypothesis testing.

   The selection of the ARL depends very much on the cost associated with a false signal and the process downtime. Selecting a smaller in-control ARL causes a higher rate of occurrence of false signals. A larger in-control ARL, on the other hand, reduces the frequency of false signals, but, at the same time it delays the detection of changes in the process mean. The cost of false signals needs to be weighted carefully before an in-control ARL is chosen.

The choice of the in-control ARL also depends on the production and sampling rates. The higher the production and sampling rates, the larger an in-control ARL can be used and vice versa.

ii) Decide on the magnitude of the minimum change in the process mean to be detected quickly. Studies by Roberts (1959) and others have shown that the EWMA chart is only better than the Shewhart chart in detecting small shifts in the process mean, but not large shifts. Therefore, the magnitude of change which we would like to detect when using the EWMA chart is normally one to two σ shifts in the process mean. This depends on how capable the process is relative to specifications and how critical the process characteristic is to the quality of the product. The cost that the company incurs for producing products when the process is not in control, e.g. rework, scrap, etc., also needs to be taken into consideration.

iii) Find the combination of λ and h that satisfies the in-control ARL in step (i) and that minimizes the size of the change selected in step (ii). This is normally done using an ARL table. ARL tables are provided by authors like Lucas and Saccucci

(1990), Crowder (1987A), Gan (1991A), etc. When using these tables, extra care is needed concerning whether the pair (λ,L) or the pair (λ,h) is given

where

λ = smoothing constant of the EWMA,

L = the control limit multiple, and

$$h = L * \sqrt{\frac{\lambda}{2-\lambda}} \qquad (1.8)$$

For example, Lucas and Saccucci (1990) and Crowder (1987A) provide tables for the ARL in terms of λ and L while Gan (1991A) gives it in terms of λ and h.

iv) Perform a sensitivity analysis by comparing the out-of-control ARL chosen in step (iii) to the out-of-control ARL for other choices of (λ,h) that produce the same in-control ARL. Select the combination of (λ,h) that produces the most desirable overall performance in term of the ARLs.

The out-of-control ARL, or the average number of samples that must be collected in order to detect a genuine process change, is, in fact, the type II error in hypothesis testing. By carrying out these four steps, we are actually trying to find a design that has acceptable type I and type II errors.

1.3 Optimal Lamda Plot

Beside using the ARL table, an optimal λ plot can also be used to select the optimal design for an EWMA control chart. The optimal λ plot is a graphical plot of the ARL table which makes the selection of λ and L as discussed in Section 1.2 easier. It is divided into 2 plots:

i) A plot of the Optimal λ vs Shift for a fixed in-control ARL (see Figure 1.1), and

ii) A plot of L vs Optimal λ for a fixed in-control ARL (see Figure 1.2).

Figure 1.1 is constructed after we have performed a sensitivity analysis for the out-of-control ARLs for different combinations of (λ,L) with a specific in-control ARL. Table 1.2 shows how this plot can be constructed. This table has been generated using the program in Appendix A.

Table 1.2: ARL table for Different Magnitudes of Shifts with an In-control ARL of 250

	λ	0.151	0.152	0.153
	L	2.656	2.657	2.659
Shifts	0.0	250.166	249.781	250.060
	0.5	27.052	27.091	27.159
	1.0	8.771	8.767	8.770
	1.5	5.051	5.045	5.041
	2.0	3.587	3.582	3.577

Table 1.3: Optimal λ for Specific Shifts in the Process Mean

	Optimal λ		
	ARL		
Shift	250	500	1000
4.0	0.91	0.89	0.86
3.5	0.84	0.8	0.76
3.0	0.73	0.68	0.62
2.5	0.58	0.52	0.46
2.0	0.41	0.36	0.32
1.5	0.27	0.24	0.22
1.0	0.152	0.134	0.118
0.5	0.055	0.047	-

	L		
	ARL		
λ	250	500	1000
1.000	2.878	3.090	3.291
0.950	2.878	3.090	3.290
0.900	2.878	3.090	3.290
0.850	2.877	3.089	3.290
0.800	2.876	3.089	3.290
0.750	2.874	3.087	3.289
0.700	2.871	3.086	3.288
0.650	2.868	3.084	3.286
0.600	2.864	3.081	3.284
0.550	2.859	3.077	3.281
0.500	2.851	3.071	3.277
0.450	2.842	3.064	3.271
0.400	2.830	3.054	3.263
0.350	2.813	3.041	3.253
0.300	2.791	3.023	3.238
0.250	2.761	2.998	3.217
0.200	2.719	2.962	3.187
0.175	2.690	2.938	3.166
0.150	2.654	2.907	3.139
0.125	2.608	2.868	3.105
0.100	2.546	2.814	3.059
0.075	2.458	2.738	2.991
0.050	2.318	2.615	2.883

Figure 1.1: Optimal λ vs Shifts

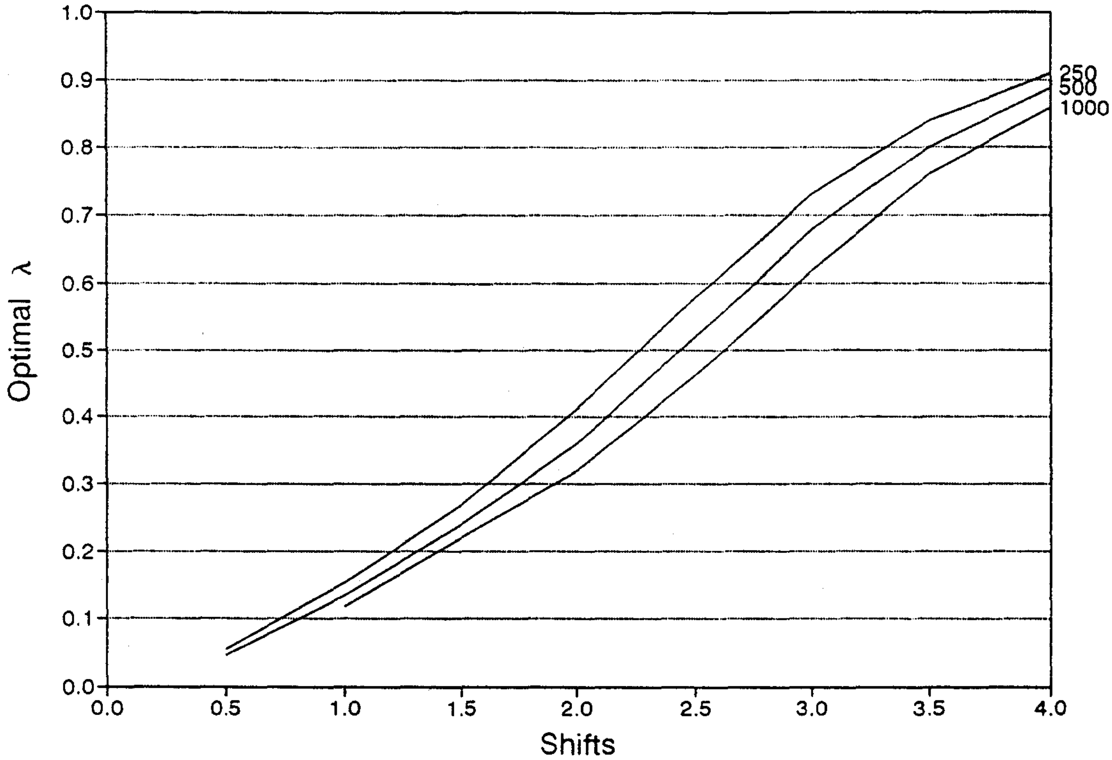
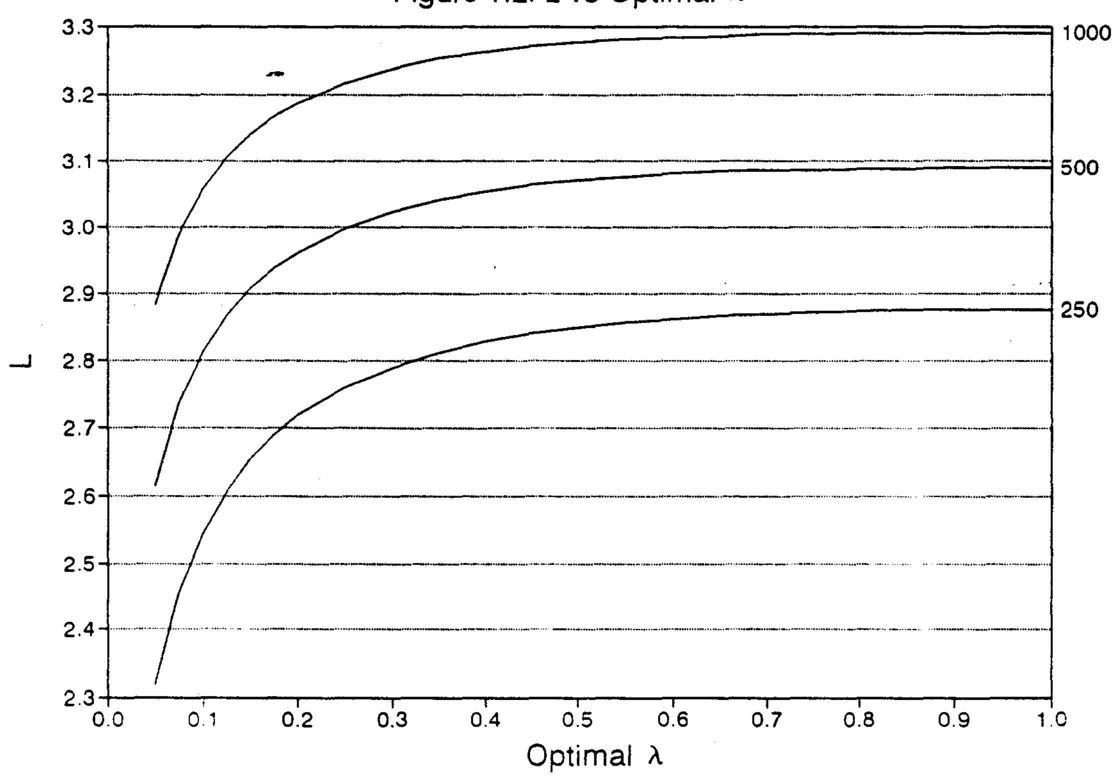


Figure 1.2: L vs Optimal λ



An optimal design should have an ARL which is as long as possible when the process is near to the target value and is short when the process has shifted to an unacceptable level. Since we normally fix the in-control ARL, we can choose the combination of λ and L that has the shortest out-of-control ARL (see steps (i) and (ii) in the four step procedure in Section 2). From Table 1.2, $\lambda = 0.152$ has the shortest ARL when the process mean has shifted by one sigma compared to all the other combinations of (λ, L) that have the same in-control ARL of 250. The optimal λ for a one sigma shift is $\lambda = 0.152$. The same method is applied to obtain the optimal λ for some other magnitude of shift. The results are plotted as shown in Figure 1.1.

The graphs in Figure 1.2 show the L corresponding to the optimal λ for specific ARLs. For example, from Table 1.2, the L for an optimal λ of 0.152 with an in-control ARL of 250 is 2.657. Other L's corresponding to various optimal λ's are obtained and plotted as in Figure 1.2.

The advantage of using an optimal λ plot is that it is easier to use. We do not have to go through the whole ARL table to find a suitable combination of λ and L when constructing an EWMA control chart. We would also be able to estimate an optimal design for a particular shift which is not found in the ARL table. For example, we might want to know the optimal design for a 1.5 sigma shift when only the

ARLs for shifts of magnitude one sigma and two sigmas are available. Estimating λ and L are easier using the optimal λ plot when compared to the ARL table because the optimal λ is not directly proportional to L.

1.4 Example

We will use an example provided by Lucas and Crosier (1982) to illustrate the EWMA control scheme. The mean of the first 10 observations is approximately zero and the mean of the last 9 observations is one. The variance is assumed to remain unchanged at the value one. We assume that the first 10 observations are collected from an in-control process which is N(0,1) distributed. The process mean is shifted by a magnitude of one sigma from observation 11 onwards. The data will also be used in the next few chapters.

Assume that the acceptable in-control ARL is chosen as 250, or $\alpha = 0.004$, after taking into consideration all the economic factors. In order to obtain an optimal design for one sigma shifts in the process mean, we can refer to the optimal λ plot in Figure 1.1. The optimal λ is found to be 0.15 and from Figure 1.2, we must choose L = 2.65 in order to have an in-control ARL of 250. The program in Appendix A can be used to find the precise design. As shown in Table 1.2, the optimal λ and L are 0.152 and 2.657.

Control limits can be calculated using

$$\mu \pm L \sqrt{\frac{\lambda}{2-\lambda}}\ \sigma_X = 0 \pm 2.657 \sqrt{\frac{0.152}{2 - 0.152}}$$
$$= \pm 0.762$$

The data together with the associated EWMA statistics are shown in Table 1.4.

Table 1.4: The EWMA statistic for λ = 0.152

t	X_t	Z_{t-1}
1	1.0	0.00
2	-0.5	0.15
3	0.0	0.05
4	-0.8	0.04
5	-0.8	-0.08
6	-1.2	-0.19
7	1.5	-0.35
8	-0.6	-0.07
9	1.0	-0.15
10	-0.9	0.03
11	1.2	-0.11
12	0.5	0.09
13	2.6	0.15
14	0.7	0.52
15	1.1	0.55
16	2.0	0.63
17	1.4	0.84 *
18	1.9	0.93 *
19	0.8	1.07 *
		1.03 *

* The EWMA statistic exceeds the control limits.

An out-of-control signal is detected within the first 16 samples ($Z_{16} > 0.762$). Since we already know that samples after the first 10 observations are shifted by one sigma, the EWMA gives an out-of-control signal in 6 observations. If a Shewhart control chart is used, we would not be able to observe any point that exceeds the Shewhart control limits in the last 9 observations.

In general, a Shewhart chart designed to have an in-control ARL of 250 can only give an out-of-control signal after an average of 33 observations (see Crowder (1989)). On the other hand, the EWMA control scheme with the same in-control ARL can detect the same shift after an average of 9 observations if $\lambda = 0.152$ is used (see Table 1.2).

CHAPTER TWO

ENHANCEMENTS OF THE EWMA FOR THE PROCESS MEAN

2.1 Introduction

The majority of the authors appear to agree that EWMA control methods should be designed based on in-control and out-of-control ARLs as discussed earlier. Many authors have proposed enhancements for the EWMA that make the EWMA control chart more effective in detecting small shifts under certain conditions. We will briefly discuss some of them. These include the following:

i) The Fast Initial Response (FIR) feature for the EWMA,

ii) The Omnibus EWMA,

iii) The EWMA for detecting gradual drifts,

iv) The EWMA for individual measurements,

v) The Multivariate EWMA, and

vi) The EWMA for monitoring a process standard deviation.

The reader is referred to the following papers for additional reading on charting methods related to the EWMA chart that are not within the scope of our discussions:

i) Rank-based EWMA for nonparametric statistics (Hackl and Ledolter (1991)),

ii) The EWMA for observations generated from a Poisson distribution (Gan (1990)),

iii) The EWMA for counts of the number of nonconformances (c) (Montgomery (1991B)),

iv) The EWMA with variable sampling intervals (Saccucci, Amin and Lucas (1992)).

and other related articles that have appeared in the Journal of Quality Technology, Technometrics, Journal of Statistical Computation and Simulation, Quality and Reliability Engineering International, etc.

To distinguish between the EWMA control scheme discussed so far and enhanced versions by other authors, we shall refer to the EWMA without any enhancements (as discussed in Chapter 1) as the EWMA from now on. Other EWMA control schemes with enhancements will be denoted by EWMA_XX, where XX represents the type of enhancement added to it.

2.2 The EWMA Control Scheme During Initial Start-Up

During the initial set-up of a control scheme, whether for a new process or for a control chart which is restarted after corrective action has been taken on an out-of-control process, the process is likely to stray away from the targeted value. Therefore, an effective control method for such a change is desirable.

Usually, the control limits are set up based on the asymptotic variance,

$$\sigma_z^2 = \left\{ \frac{\lambda}{2-\lambda} \right\} \sigma_x^2 \qquad \left[\text{see Chapter 1, equation 1.7}\right]$$

However, as mentioned in Chapter 1, Section 1, unless λ is small, the variance would converge to its asymptotic value fairly quickly. Unfortunately, we are using the EWMA to monitor small shifts in the process mean. This means that the λ that we are using is small; normally, it ranges from 0.1 to 0.3. The variance would converge slowly to its asymptotic value. Therefore, it would not be very sensitive to any abnormality that may occur during the initial phase of the control chart.

One suggestion is to extend the Fast Initial Response (FIR) feature for the CUSUM chart to the EWMA chart (Lucas and Saccucci (1990)). Normally, we would have only one EWMA where the target mean or actual mean is used as Z_0. If FIR is used, another two more EWMAs, each with different starting values or Head Starts (HS) are used. One of the starting values is above the target value, viz., Z_0^+, while the other one is below the target value, viz., Z_0^-. We shall refer to this type of EWMA as the EWMA with head start, EWMA_HS. What we mean by EWMA_HS is that, instead of using the target value or the actual mean as Z_0, Z_0 is chosen from a value between the process mean and the control limit. We can define Z_0 as

$$Z_0^{\pm} = \mu \pm \theta * \sigma_Z , \quad \text{where } 0 \le \theta < L ,$$

$$\text{or} \quad Z_0^{\pm} = \mu \pm \frac{\theta}{L} \left\{ \frac{\text{UCL}-\text{LCL}}{2} \right\} , \tag{2.1}$$

where $0 \le \theta < L$,

μ = process mean,

L = control limit multiple, and

UCL and LCL are the upper and lower control limits, respectively.

For example, if we have decided to use an EWMA control chart with center line 1, upper control limit 2.1 and lower control limit -0.1, and we have decided to use 50% HS, then

$$Z_0^{+} = 1 + 0.5 \left\{ \frac{2.1-(-0.1)}{2} \right\} = 1.55$$

$$Z_0^{-} = 1 - 0.5 \left\{ \frac{2.1-(-0.1)}{2} \right\} = 0.45.$$

The EWMA_HS normally gives an out-of-control signal faster if the process mean shifts during the initial set-up. The two EWMA with HS tend to converge if there is no shift in the process mean. How fast they converge again depends on the λ chosen. These two EWMAs with head start can be stopped if they are sufficiently close to the normal EWMA which uses either the target mean or the actual mean as Z_0. This method, however, will become a burden if there are a lot of process variables to monitor.

MacGregor and Harris (1990) have proposed using the exact variance to construct the control limits instead of using FIR during initial set-up. When using the method suggested by MacGregor and Harris (1990), the exact variance at time t,

$$\sigma_z^2 = \left\{ \frac{\lambda}{2-\lambda} * \left[1-(1-\lambda)^{2t}\right] \right\} \sigma_x^2 \quad , \qquad \left[\text{see Chapter 1, equation 1.6}\right]$$

is used instead of the asymptotic variance

$$\sigma_z^2 = \left\{ \frac{\lambda}{2-\lambda} \right\} \sigma_x^2 \quad , \qquad \left[\text{see Chapter 1, equation 1.7}\right]$$

to construct the control limits. They have also proved using simulation that this method, which we shall call EWMA_EV, gives similar ARLs when compared to EWMA_HS (see Table 2.1).

Table 2.1: Comparison of ARLs for EWMA_HS and EWMA_EV

$\lambda=0.133$, $L=2.856$

Shift	EWMA_EV	EWMA_HS
0	452	434
0.5σ	30	27
1.0σ	8.4	7
2.0σ	2.7	2.6

Source: Adapted from MacGregor, J.F. and Harris, T.J. (1990).

Besides having the ability to similarly detect any abnormality during initial set-up, it is also not neccessary in the EWMA_EV to plot two more additional EWMAs with Z_0^+ and Z_0^- as starting values as in the EWMA_HS.

An example of an EWMA chart plotted using the exact variance to set up the control limits is the following:

Table 2.2: EWMA Using the Exact Variance

$\lambda = 0.152$

t	X_t	Z_t	UCL	LCL
		0.00		
1	1.0	0.15	0.40	-0.40
2	-0.5	0.05	0.53	-0.53
3	0.0	0.04	0.60	-0.60
4	-0.8	-0.08	0.65	-0.65
5	-0.8	-0.19	0.68	-0.68
6	-1.2	-0.35	0.71	-0.71
7	1.5	-0.07	0.72	-0.72
8	-0.6	-0.15	0.73	-0.73
9	1.0	0.03	0.74	-0.74
10	-0.9	-0.11	0.75	-0.75
11	1.2	0.09	0.75	-0.75
12	0.5	0.15	0.75	-0.75
13	2.6	0.52	0.76	-0.76
14	0.7	0.55	0.76	-0.76
15	1.1	0.63	0.76	-0.76
16	2.0	0.84	0.76	-0.76
17	1.4	0.93	0.76	-0.76
18	1.9	1.07	0.76	-0.76
19	0.8	1.03	0.76	-0.76

The EWMA chart using the exact variance to calculate the control limits is plotted as shown in Figure 2.1.