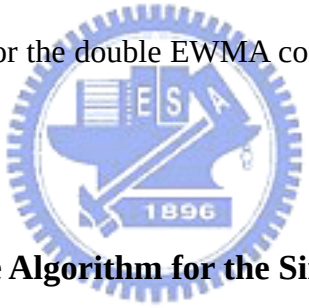
CHAPTER 3

PROPOSED EWMA CONTROLLERS

The performance of the EWMA controlled process is based on choosing the correct EWMA gain. Most related researches have focused on analyzing the optimal EWMA gain in the static condition. Unfortunately, a process environment is usually dynamic in a real manufacturing world. In order to achieve a better performance in the dynamic system, developing a method of on-line tuning of the EWMA controller parameters is an important issue. The main objectives of this chapter are: (1) Develop a neural network (NN) based adaptive algorithm for the single EWMA controller. (2) Enhance the efficiency of the neural-based adaptive algorithm. (3) Develop a time-varying tuning strategy for the double EWMA controller. (4) Develop dynamical double EWMA controller.



3.1 Neural-Based Adaptive Algorithm for the Single EWMA Controller

A methodology was developed under the framework of neural networks to conduct on-line tuning of the parameters of the EWMA controller. The input feature of the neural structure is the sample autocorrelation function (SACF) and the output unit is an estimator of the EWMA controller parameter at run *t*. The theoretical autocorrelation function (ACF) at lag h is defined as follows:

$$\rho(h) = \frac{Cov(y_t, y_{t+h})}{\sigma(y_t)\sigma(y_{t+h})} \tag{3.1}$$

Equation (3.1) can be estimated by the SACF:

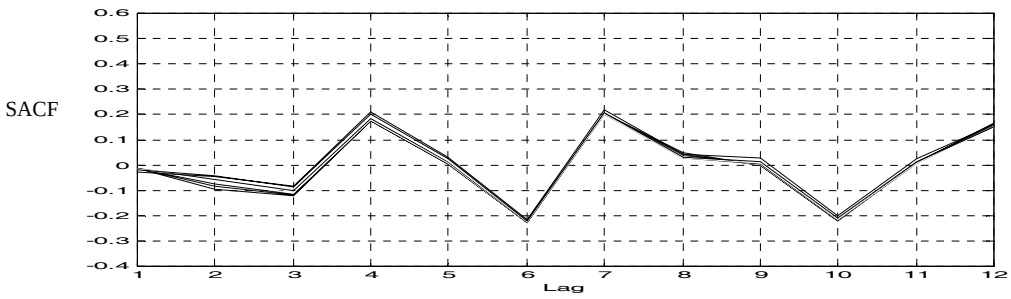
$$\hat{\rho}(h) = \frac{\gamma_h}{\gamma_0} \tag{3.2}$$

where $\gamma_h = \dfrac{\sum_{i=1}^{n-h}(y_i - \bar{y})(y_{i+h} - \bar{y})}{n}$, n is the sample size, and $\bar{y}$ is the sample mean of

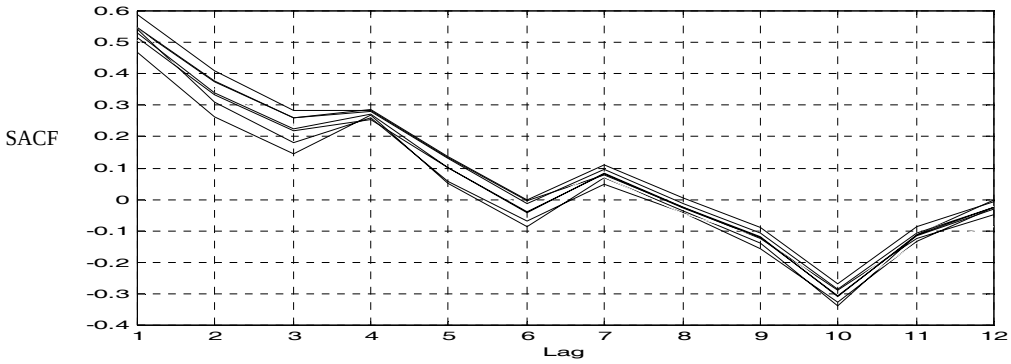
process output.

The idea of selecting the SACF to be the input feature in the neural network can be described simply in Figure 3.1. Figure 3.1(a) is the family of SACF patterns given that the controller parameter is 0.1, and Figure 3.1(b) is the family of SACF patterns given that the controller parameter is 0.9. On the one hand, we can see that the SACF behaves [in Figure 3.1(b)] like the exponential decay, which implies that the more non-stationary the process is, the larger the value of the controller parameter will be in order to compensate the process. On the other hand, the tendency of the SACF behaves similarly when the controller parameter has a specific value (say $\lambda = 0.1$ or 0.9). So, our objective is to estimate the controller parameter through the tendency of the SACF pattern.

The structure of the proposed adaptive neural-based EWMA controller is shown in Figure 3.2. At first the controlled process output was sent to the SACF block to calculate the $\hat{\rho}(h)$, and then the SACF pattern over lag h was fed into the Trained NN Model block to estimate the controller parameter. After estimating the parameter, we updated the EWMA controller parameter with time to provide a better control performance.



(a)



(b)

Figure 3.1 The parameter with (a) $\lambda = 0.1$ (b) $\lambda = 0.9$

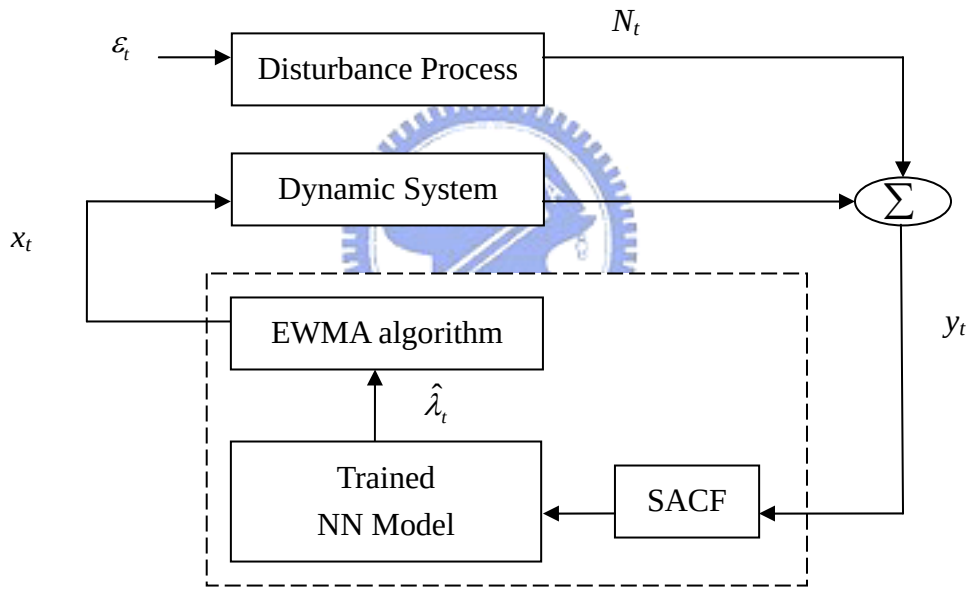


Figure 3.2 NN-based EWMA controller

3.2 Enhanced Neural Adaptive Algorithm for the Single EWMA Controller

In this section, our objective is to reduce the learning speed of the neural network in order to enhance the neural-based adaptive EWMA controller. The input features of the neural structure we selected are sample auto-correlation function (SACF) and sample partial auto-correlation function (SPACF). The output unit is an estimate of

the EWMA controller parameter at run *t*. The theoretical partial auto-correlation function (PACF) is defined as follows:

$$\rho_{hh} = \frac{Cov(y_t, y_{t+h} | y_{t+1}, y_{t+2}, ..., y_{t+h-1})}{\sigma(y_t)\sigma(y_{t+h})} \tag{3.3}$$

which can be estimated by the SPACF as follows:

$$\hat{\rho}_{hh} = \frac{\hat{\rho}_h - \sum_{j=1}^{h-1} \hat{\rho}_{h-1,j}\hat{\rho}_{h-j}}{1 - \sum_{j=1}^{h-1} \hat{\rho}_{h-1,j}\hat{\rho}_j}, \qquad h = 3, ... , \infty \tag{3.4}$$

where

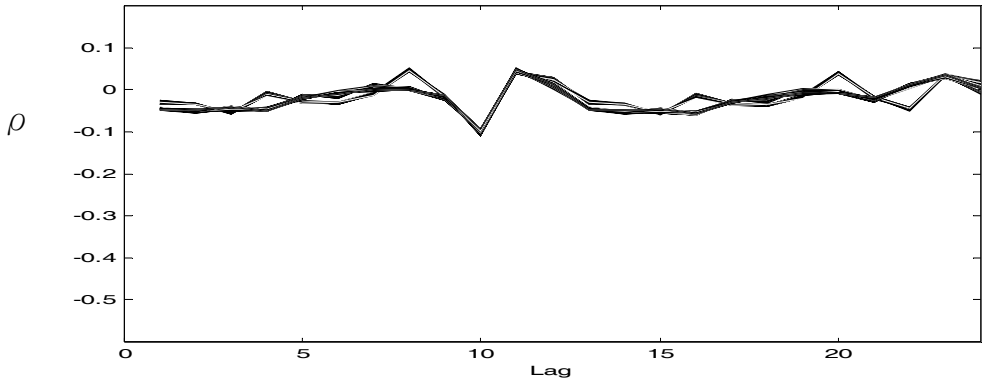
$$\hat{\rho}_{hj} = \hat{\rho}_{h-1,j} - \hat{\rho}_{hh}\hat{\rho}_{h-1,h-j} \qquad h = 2, ..., \infty; \quad j = 1,2, ..., h-1 \tag{3.5}$$

and $\hat{\rho}_{11} = \hat{\rho}_1$, $\hat{\rho}_{22} = \frac{\hat{\rho}_2 - \hat{\rho}_1^2}{1 - \hat{\rho}_1^2}$.

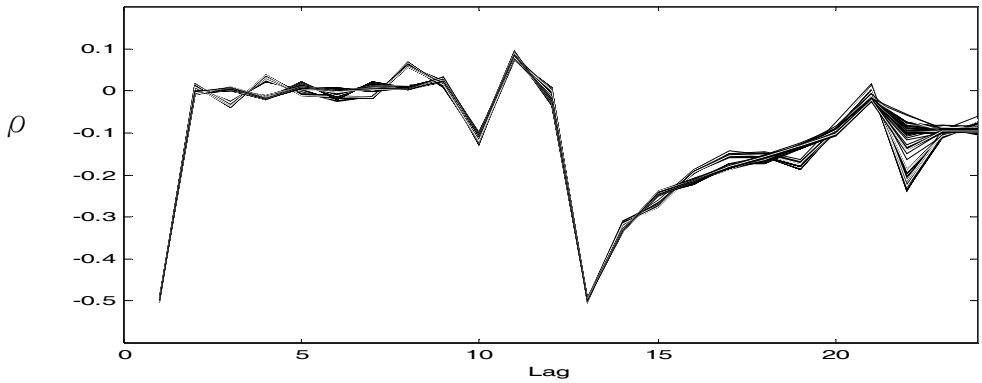
Opinions of selecting SACF and SPACF statistics to be input features can be described simply from Figure 3.3. Figure 3.3(a) shows the family of SACF/SPACF (denoted ρ) patterns under the condition of a perfect controlled process up to run *t*, given the disturbance model obeys Equation (2.8) and the controller parameter follows Equation (2.10). If we receive this type of SACF/SPACF patterns, then we will set the controller parameter at next run (*t+1*) to be the same with previous run. Figure 3.3(b) shows the simulated family of SACF/SPACF patterns with parameter $\theta = 1$ (white noise process) in Equation (2.8) and the full adjustment ($\lambda = 1$) in Equation (2.10) up to run *t*. Obviously, Figure 3.3(b) behaves more non-stationary than Figure 3.3(a), because of the incorrect choosing the controller parameter. If the neural network receives the types of patterns such as shown in Figure 3.3(b) at run *t*, then it will respond by setting the controller parameter to be zero at the next run to meet the optimal condition. So, as per the above, our objective is to estimate the

controller parameter adaptively through pattern recognition on the SACF and SPACF patterns.

The structure of the proposed enhanced neural network adaptive single EWMA controller is shown in Figure 3.4. At first, the controlled quality characteristic was sent to SACF and SPACF blocks to calculate the statistic individually. The combined (denoted as the black bar) SACF/SPACF pattern was then sent to the Trained NN Model block to estimate the controller parameter for the next run. After estimating the parameter, we should update the single EWMA controller parameter dynamically to provide a better control performance.



(a)



(b)

Figure 3.3 The family of SACF/SPACF patterns (a) perfect controlled (b) $\theta = 1; \lambda = 1$

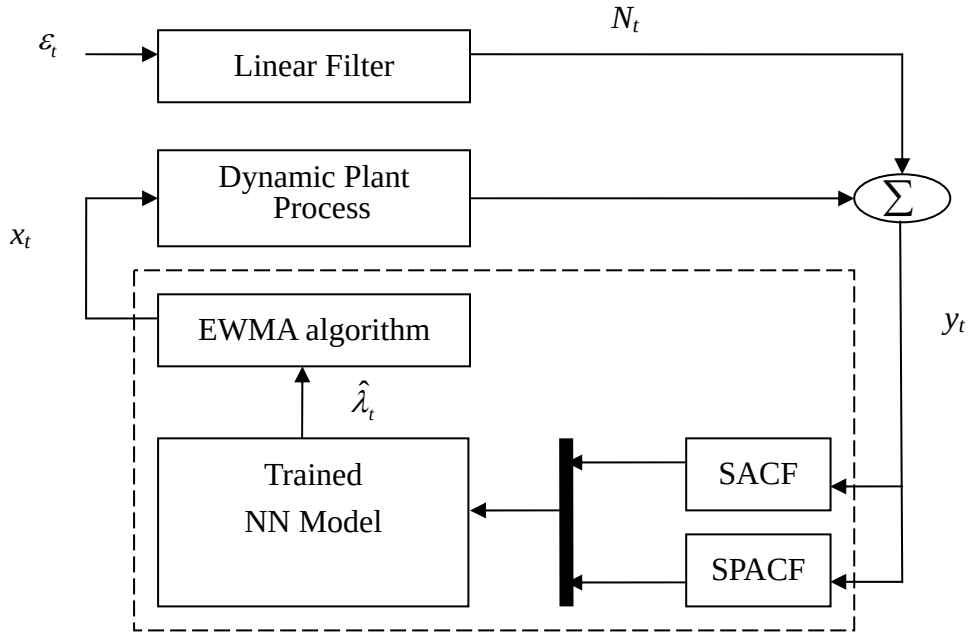


Figure 3.4 Enhanced NN adaptive single EWMA controller

3.3 A Time-Varying Weights Tuning Strategy for the Double EWMA Controller

In order to enhance the performance of the double EWMA controller, a simple but effective time-varying weights tuning algorithm will be proposed in this section. We will first present a preliminary model of the time-varying control scheme and then modify it to be our proposed tuning method.

It was intuition to initially use the all-bias solution to bring the process on target, and then use the all-variance solution to reduce the process oscillations around the target. However, it has shown to be an inefficient tuning method from the viewpoint of “abrupt change”. Therefore, we are attempting to use the viewpoint of “gradual change”, which means using higher weights first and then “gradually” reducing them to the all-variance solution weights. We call this control strategy a GC (gradual change) control scheme, and it can be expressed as follows:

$$\lambda_1(t) = \lambda_{1,v} + (f_1)^t \tag{3.6}$$

$$\lambda_2(t) = \lambda_{2,v} + (f_2)^t \tag{3.7}$$

where f_1 $(0 \le f_1 < 1)$ and f_2 $(0 \le f_2 < 1)$ denote the discount factor. $\lambda_{1,v}$ and $\lambda_{2,v}$ individually represent the c_t's and D_t's all-variance solution weight. From Equations (3.6) and (3.7), we can see that $\lambda_1(t)$ and $\lambda_2(t)$ approach the all-variance solution as time approaches to infinity.

However, there is a problem for adding a discount factor in Equation (3.6). Consider a drifting process in Equation (2.13). Figure 3.5 shows the results of the controlled process (without noise term) when the discount value (f_1) is large, moderate and 0. We can see that a large discount value makes it difficult for the process to converge to the target value. On the contrary, when the discount value is 0, the process converges quickly to the target. Therefore, we are tempering it by setting the discount value in Equation (3.6) to be 0. That is:

$$\lambda_1(t) = \lambda_{1,v} \tag{3.8}$$

$$\lambda_2(t) = \lambda_{2,v} + (f)^t \tag{3.9}$$

where f $(0 \le f < 1)$ expresses the discount factor. We call Equations (3.8) and (3.9) the MGC (modified gradual change) control scheme.

From Equation (2.30), it can be shown that if the double EWMA controller with the fixed weights control scheme, then the following equation holds as:

$$(\lambda_1, \lambda_2) = (\max\{\lambda_1, \lambda_2\}, \min\{\lambda_1, \lambda_2\}) = (\min\{\lambda_1, \lambda_2\}, \max\{\lambda_1, \lambda_2\}) \tag{3.10}$$

But, for the time-varying weights, the above equation does not hold. Therefore, there are two cases in the MGC control scheme, they are MGC-1:

$$\lambda_1(t) = \min\{\lambda_{1,v}, \lambda_{2,v}\} \tag{3.11}$$

$$\lambda_2(t) = \max\{\lambda_{1,v}, \lambda_{2,v}\} + (f)^t \tag{3.12}$$

and MGC-2:

$$\lambda_1(t) = \max\{\lambda_{1,v}, \lambda_{2,v}\} \tag{3.13}$$

$$\lambda_2(t) = \min\{\lambda_{1,v}, \lambda_{2,v}\} + (f)^t \tag{3.14}$$

In order to compare the tuning method between MGC-1 and MGC-2, we assume the discount factor *f* to be fixed as a constant. Figure 3.6 shows the controlled process under MGC-1 versus MGC-2 control scheme. It shows that both control schemes converge to the target at almost the same time, but the controlled output under the MGC-2 tuning method shows much smaller offsets from the target. Thus, we will adopt the MGC-2 to be our proposed time-varying weighs tuning method for the EWMA controller.

The advantage of adding a discount factor in our proposed tuning method is the quick response to the initial transient effect. In Figure 3.7, the dash and solid line individually represent the controlled output 'with' and 'without' adding the discount value. For the case of $f \neq 0$, the double EWMA controller compensates for the initial transient effect more quickly than in the case of $f = 0$. From the proposed tuning equations, we know that the performance of the controlled process output depends on setting the discount factor (*f*). Even though a larger discount value can quickly compensate for the initial transient effect, it may cause oscillations. Therefore, we will present how to determine the discount parameter in the Section 5.1.

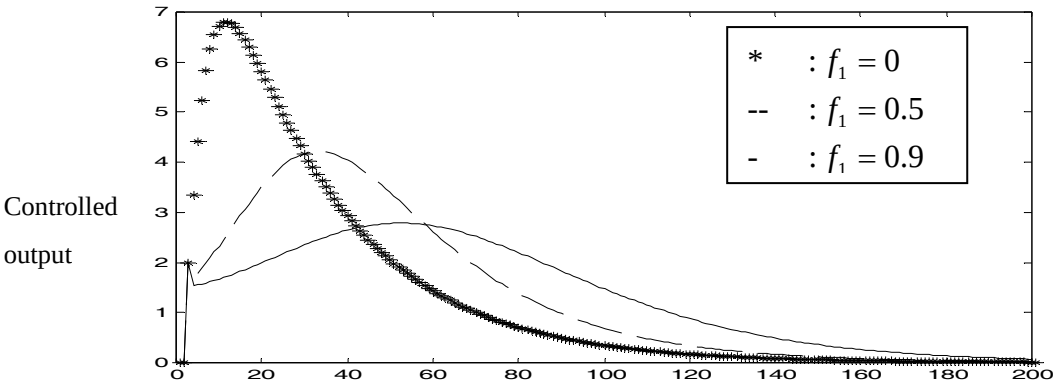


Figure 3.5 Controlled output with f_1 being large, moderate and

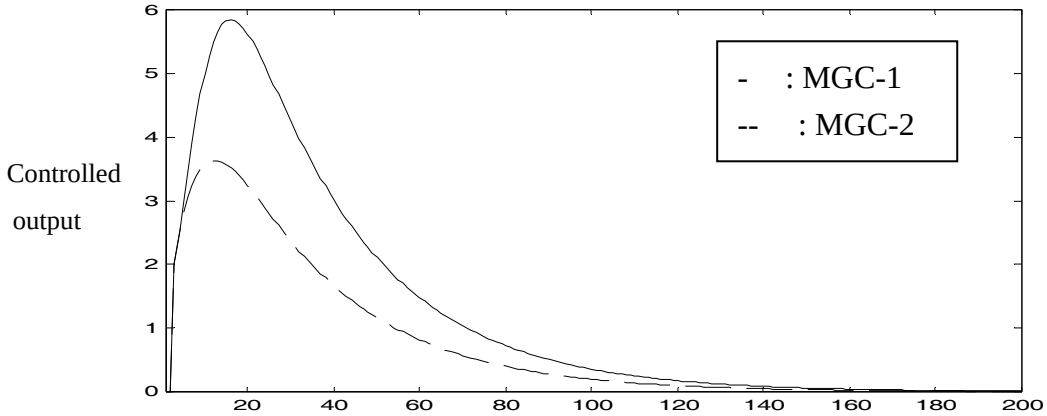


Figure 3.6 MGC-1 v.s. MGC-2 tuning method

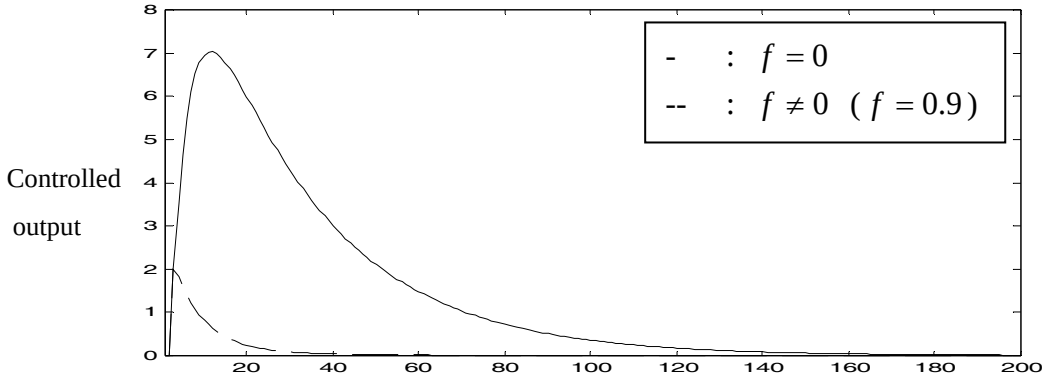


Figure 3.7 Controlled output with and without discount

3.4 Dynamic Double EWMA Controller Under Wear-Output Process

Del Castillo and Hurwitz (1997) used the recursive least squares (RLS) theory to continuously estimate the single EWMA parameter. Patel and Jenkins (2000) proposed an adaptive single EWMA control algorithm by using the signal-to-noise ratio (SNR).

As mentioned in the literature, their objective was to develop a dynamic tuning method for the single EWMA controller. Relevant research regarding the dynamic tuning, double EWMA controller is seldom mentioned in the literature. Therefore, we will develop a dynamic tuning double EWMA feedback controller under a wear-out process in this section. The proposed controller can dynamically change the control parameters in response to disturbance changes. Figure 3.8 shows the architecture of the proposed dynamic tuning, double EWMA controller. We can observe that the proposed controller contains two main modules, and they are: the 'Trigger' and the 'Dynamic Tuning Loop' modules.

Before taking a control action, we have to find any shift that the process may have undergone. Therefore, the objective of the Trigger module is to detect any process shift, and then to reset the double EWMA control parameters. The widely used EWMA technique in statistical control can be adopted here (Hunter 1986, Montgomery 1996, Wang and Mahajan 1996, Guo and Chen 2002). The EWMA of the process output can be calculated as

$$z_t = qy_t + (1-q)z_{t-1} \tag{3.15}$$

where q is a weight factor of the EWMA. If the sample number is moderately large, then the standard deviation of z converges to its asymptotic value as

$$\sigma_z = \sqrt{\frac{q}{2-q}}\sigma_y \tag{3.16}$$

Thus, the control action limit can be set to $L\sigma_z$. By taking $L = 3$ (the usual 3-sigma

limits), and selecting $q = 0.2$, the control limit simply becomes σ_y.

When the control chart triggers a signal that implies a change in the disturbance, the objective of the triggered Dynamic Tuning Loop module is to compensate for the new disturbance. From Figure 3.9, consider the change in the disturbance at run t' (unknown) while the control chart detects the shift at run *k*. The Dynamic Tuning Loop module can be expressed as follows:

$$\begin{aligned} \lambda_1(t) &= \max\{\lambda_{1,v}, \lambda_{2,v}\} \\ \lambda_2(t) &= \min\{\lambda_{1,v}, \lambda_{2,v}\} + (f)^{d+(t-k)} \end{aligned} \tag{3.17}$$

where *d* denotes the time delay between the point of change and the point of detection. Thus, *d* can roughly be estimated by the average run length, that is $d \approx ARL_1$. Table 1 shows the ARL_0 (in-control ARL) and ARL_1 (out-of-control ARL) for the Shewhart and EWMA control chart under several drifting rates. The obtained results are calculated using Monte Carlo simulations. In Table 3.1, the EWMA chart is shown to be more sensitive than the Shewhart chart for detecting slow to moderate drifting rates (say $0.1 \le \delta \le 0.5$). So, the early detection is advantageous, since it allows control action to be taken in advance of any potentially bad runs.

To judge if the Dynamic Tuning Loop is completed or not, we can use the following criterion:

$$\lambda_2(t) = \begin{cases} \max\{\min(\lambda_{1,v}, \lambda_{2,v}), \min(\lambda_{1,v}, \lambda_{2,v}) + (f)^{ARL_1+(t-k)}\} & \textit{if not completed} \\ \min\{\lambda_{1,v}, \lambda_{2,v}\} & \textit{if completed} \end{cases}$$

In Figure 3.8, if the tuning loop shows completed and the EWMA chart triggers a new signal, then this implies that the disturbances parameter (δ) changed. Like Equation (2.27), the asymptotically estimate of δ under MGC-2 tuning strategy is $D_t \max\{\lambda_{1,v}, \lambda_{2,v}\}$. When δ has been estimated, we can resolve Equation (2.33) and obtain a new all-variance solution weights. After that, the Dynamic Tuning Loop

module can be triggered to adjust the double EWMA controller weights.The efficiency of the proposed dynamic tuning double EWMA controller will be illustrated in Section 5.3.

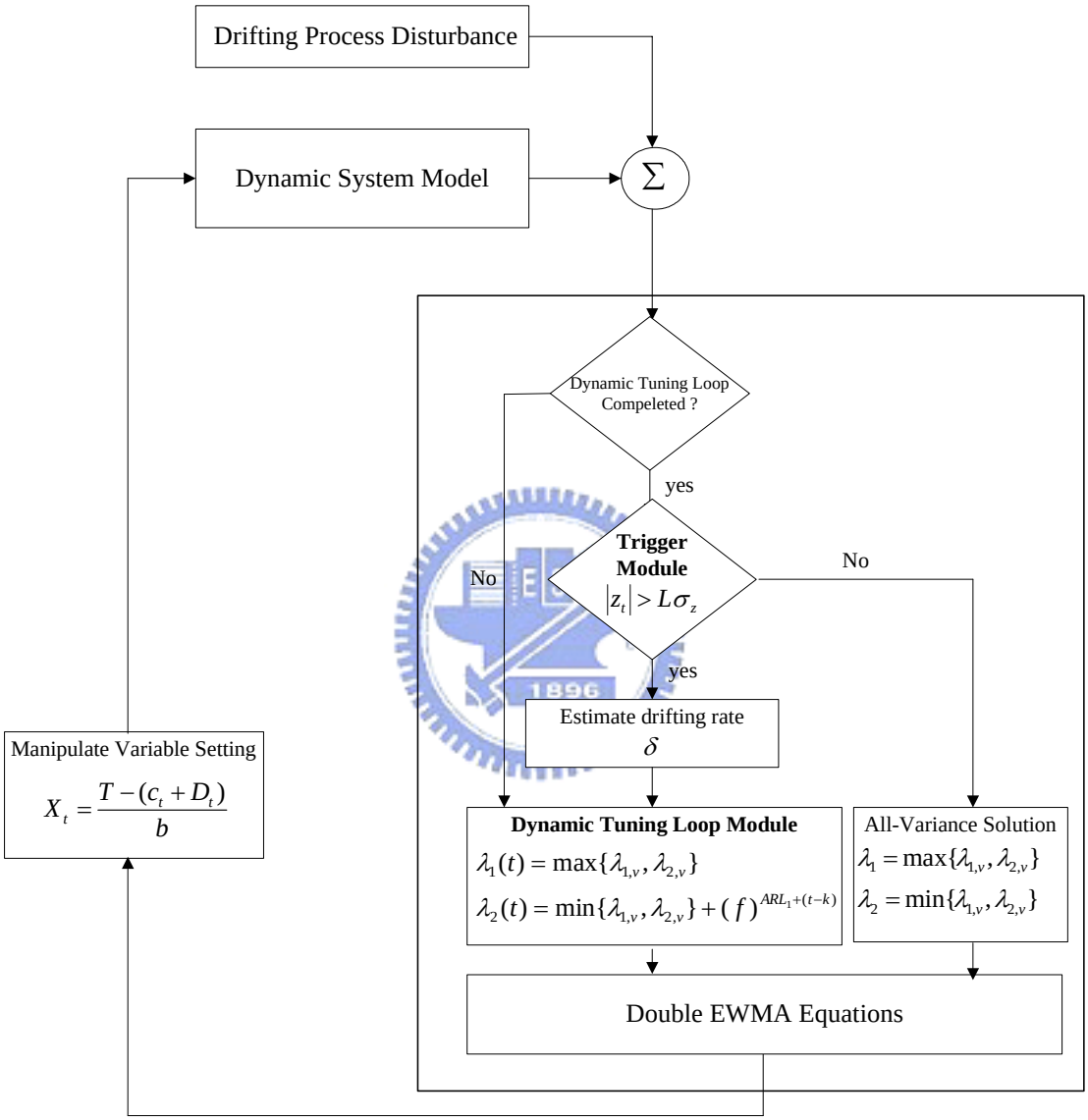


Figure 3.8 Dynamic tuning double EWMA Controller under wear-out

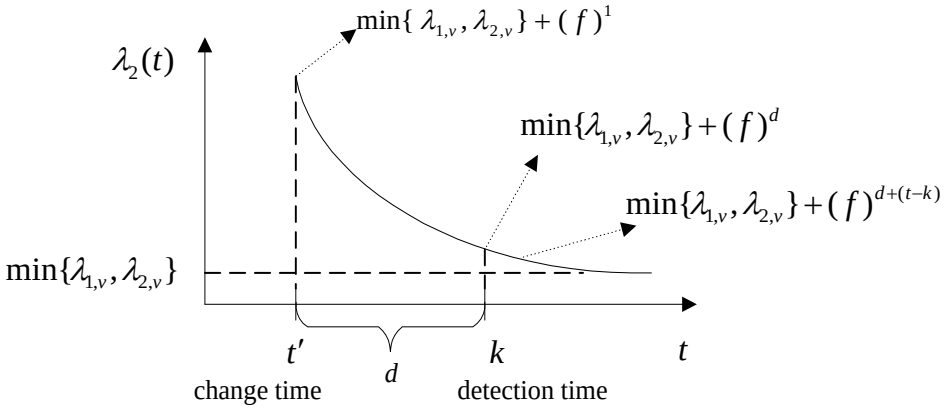


Figure 3.9 The trigged tuning loop (i.e. $\lambda_2(t)$)

Table 3.1 ARL for Shewhart and EWMA control chart

δ (σ /run)	Shewhart $L = 3.1$	EWMA $q = 0.2,\ L = 2.962$
0	500	500
0.1	20.3530	14.203
0.2	11.8530	9.5033
0.3	8.6167	7.7300
0.4	6.8867	6.6567
0.5	5.6800	5.5233
0.6	4.9667	5.5100
0.7	4.4800	5.1367
0.8	4.0433	4.8433
0.9	3.7067	4.6000
1.0	3.4633	4.3700