Statistical process control for serially correlated data
Wieringa, Jakob
Chapter 4

EWMA-type control charts for the mean

In the previous chapter, we discussed Shewhart-type control charts. These control charts only use the current observation or sample to monitor the process. In the next two chapters we consider control charts that also utilize previous observations. In Chapter 5, the *Cumulative SUM* (CUSUM) chart is discussed. In its basic form, an unweighted cumulative sum of the (standardized) observations is plotted against time in the CUSUM chart. This chart has a long ‘memory’.

In the present chapter, we consider the *Exponentially Weighted Moving Average* (EWMA) control chart. Like the CUSUM, the EWMA utilizes all previous observations, but the weight attached to data is exponentially declining as the observations get older and older. By varying the parameter of the EWMA statistic the ‘memory’ of the EWMA control chart can be influenced. A control chart based on the EWMA was introduced by Roberts (1959). More recent references include Hunter (1986), Crowder (1987), and Lucas and Saccucci (1990).

In the previous chapter, the EWMA statistic was used as a local estimator for the level of the data. The EWMA ‘smoothes out’ the effect of single disturbances, and shows the behavior of the level of the data. This suggests using the EWMA as a statistic to monitor the mean of a process. Originally, the EWMA was developed by time series analysts to distinguish short term variation from long term variation such as trends and cyclic behavior.

Another application of the EWMA was mentioned in Section 3.7 of the previous chapter. The EWMA of previous observations provides an
approximate one-step-ahead predictor. It is easily shown that the EWMA is an optimal (in the sense of minimal Mean Squared Error (MSE)) one-step-ahead predictor if the underlying time series model is IMA(1,1). This property will be illustrated in Chapter 8.

The setup of this chapter is similar to that of the previous chapter. In Section 4.1, we discuss the EWMA control chart for independent observations. The ARL curve that is derived in this section will serve as a benchmark in Sections 4.3, 4.4 and 4.5, where the modified EWMA chart, the EWMA chart of residuals, and the EWMA of modified residuals will be discussed, respectively. In Section 4.6, the ARL behavior of the three types of EWMA control charts are compared, and conclusions are presented.

### 4.1 The EWMA control chart for i.i.d. observations

As in the previous chapter, the sequence of independent observations of a quality characteristic of interest is denoted by \{X_t\}. Assume that \(X_t\), an observation at time \(t\) with \(t \in \mathbb{Z}\), are independently distributed as

\[
X_t \sim \mathcal{N}(\mu_t, \sigma_X^2)
\]

for \(t \in \mathbb{Z}\),

where the index \(t\) of \(\mu_t\) indicates that the mean of the observations may shift over time. The value of the EWMA statistic at time \(t\), which we will denote by \(W_{X,t}\), is computed as follows

\[
W_{X,t} = \lambda X_t + (1 - \lambda)W_{X,t-1},
\]

where the parameter \(\lambda\) is a constant satisfying \(\lambda \in (0, 1)\). Usually \(W_{X,0}\) is set equal to a target value, or (an estimation of) the mean \(\mu\). All information on previous observations that is needed for computing \(W_{X,t}\) is stored in \(W_{X,t-1}\).

If \(W_{X,t}\) is interpreted as a one-step-ahead predictor of \(X_{t+1}\), formula (4.1) can be rewritten as

\[
\hat{X}_{t+1} = \hat{X}_t + \lambda(X_t - \hat{X}_t).
\]

This formula shows that the predictor for \(X_{t+1}\) equals the predictor for \(X_t\), corrected with a fraction of the error made in the forecast of \(X_t\). As
4.1. THE EWMA CONTROL CHART FOR I.I.D. OBSERVATIONS

Hunter (1986) remarks, this makes plotting of the EWMA almost as easy as plotting the successive observations.

Equation (4.1) can also be rewritten as

\[ W_{X,t} = \begin{cases} W_{X,0} & \text{for } t = 0 \\ \lambda \sum_{i=0}^{t-1} (1 - \lambda)^i X_{t-i} + (1 - \lambda)^t W_{X,0} & \text{for } t = 1, 2, \ldots \end{cases} \] (4.2)

From expression (4.2) it is clear why the EWMA is sometimes called a geometric moving average, since the weights of past observations are declining as in a geometric series. The use of the term average can be justified by observing that for any \( t \) the weights sum to one, since

\[ \lambda \sum_{i=0}^{t-1} (1 - \lambda)^i + (1 - \lambda)^t = 1. \]

The choice of \( \lambda \) determines the decline of the weights, and thereby the effect of past observations in the computation of \( W_{X,t} \). However, for all possible choices of \( \lambda \), more recent observations always receive more weight in the computation of \( W_{X,t} \) than older observations. If \( \lambda \to 1 \) then \( W_{X,t} \to X_t \), and the EWMA places all of its weight on the most recent observation. The EWMA control chart will then behave as the Shewhart control chart. If \( \lambda \to 0 \), then the most recent observation receives a small weight, whereas the weight attached to previous observations only slightly declines with the age of the observations. The EWMA then takes on the appearance of the CUSUM. How to determine \( \lambda \) is discussed later in this section.

If we take \( W_{X,0} = \mu \) it is easily seen from Equation (4.2) that the expectation of \( W_{X,t} \) is equal to \( \mu \) whereas for the variance of \( W_{X,t} \) we have

\[ \sigma_{W_{X,t}}^2 = \sigma_X^2 \left( \frac{\lambda}{2 - \lambda} \right) \left[ 1 - (1 - \lambda)^{2t} \right]. \] (4.3)

Using this expression, the control limits of an EWMA chart for the mean of \( \{X_t\} \) can be computed. The lower control limit (UCL) at time \( t \) is constructed as follows

\[ \text{LCL}_t = \mu - c\sigma_X \sqrt{\left( \frac{\lambda}{2 - \lambda} \right) \left[ 1 - (1 - \lambda)^{2t} \right]}, \] (4.4)
while the upper control limit (UCL) of the EWMA is computed as

\[
UCL_t = \mu + c\sigma_X \sqrt{\left(\frac{\lambda}{2-\lambda}\right)}[1 - (1 - \lambda)^{2t}] 
\]  

(4.5)

The EWMA chart generates an out-of-control signal at time \( t \) if the realization of \( W_{X,t} \) at time \( t \) (which we will denote by \( w_t \)) is larger than \( UCL_t \) or smaller than \( LCL_t \). In Equations (4.4) and (4.5) \( c > 0 \) is a constant that needs to be chosen by the designer of the control chart.

From Equation (4.3) we can see that \( \sigma_{W_{X,t}}^2 \) is increasing over time. Hence, the control limits in Equations (4.4) and (4.5) will become wider. However, unless \( \lambda \) is very small, \( \sigma_{W_{X,t}}^2 \) converges very quickly to \( \lambda/(2-\lambda)\sigma_X^2 \). When constant limits are preferred, the computations will be based on the asymptotic standard deviation instead of the exact \( \sigma_{W_{X,t}} \). Control limits are then obtained by

\[
LCL = \mu - c\sigma_X \sqrt{\left(\frac{\lambda}{2-\lambda}\right)} 
\]  

and

\[
UCL = \mu + c\sigma_X \sqrt{\left(\frac{\lambda}{2-\lambda}\right)} 
\]  

(4.6) (4.7)

Usually, \( \mu \) and \( \sigma_X \) will not be known in practice. Therefore, \( \mu \) and \( \sigma_X \) will be replaced by estimators \( \hat{\mu} \) and \( \hat{\sigma}_X \) in Equations (4.4) through (4.7).

The EWMA control chart has two parameters, \( c \) and \( \lambda \). Their value will be set based on requirements on the ARL curve. The first to study the ARL of the EWMA was Roberts (1959). Using simulation results, he derived nomograms for the ARL of normally distributed variables. Numerical results were obtained by Robinson and Ho (1978), they evaluated the ARL of the EWMA control chart using Edgeworth expansion. Crowder (1987) showed that the ARL of the EWMA chart can be written as a so-called Fredholm integral of the second kind. The ARL curve can then be evaluated by approximating the integral numerically, see Appendix B. This Fredholm-integral approach of Crowder will be discussed in this chapter. Lucas and Saccucci (1990) used a Markov-chain approach to evaluate the ARL of the EWMA control chart. This approach is discussed in Appendix C.
4.1. THE EWMA CONTROL CHART FOR I.I.D. OBSERVATIONS

Assume that an out-of-control signal is given if $|W_{X,t} - \mu| > h$ for some constant $h$, and define $L_{W_X}(\delta, u)$ as the ARL of the EWMA chart for the mean of $\{X_t\}$, given that the shift in the mean is equal to $\delta \sigma_X$, and that the EWMA starts in $W_{X,0} = u$. The run length is 1 if $x_1$, a realization of $X_1$, is such that $|(1 - \lambda)u + \lambda x_1 - \mu| > h$. Otherwise the run continues from $(1 - \lambda)u + \lambda x_1 - \mu$. From this point on, we expect an additional run length of $L_{W_X}(\delta, (1 - \lambda)u + \lambda x_1 - \mu)$. This leads to the following integral equation for $L_{W_X}(\delta, \cdot)$

$$L_{W_X}(\delta, u) = 1 \cdot \Pr((|1 - \lambda)u + \lambda X_1 - \mu| > h) +$$

$$+ \int_{|(1 - \lambda)u + \lambda x_1 - \mu| \leq h} [1 + L_{W_X}(\delta, (1 - \lambda)u + \lambda x_1 - \mu)] f(x_1) dx_1$$

$$= 1 + \frac{1}{\lambda} \int_{-h}^{h} L_{W_X}(\delta, x) f\left(\frac{x + (1 - \lambda)u}{\lambda}\right) dx,$$  \hspace{1cm} (4.8)

where $f(\cdot)$ denotes the probability density function of $X_t$ ($t = 1, 2, \cdots$). Equation (4.8) is a Fredholm integral of the second kind. Some considerations on Fredholm-integral equations of the second kind are discussed in Appendix B.

The ARL of the EWMA chart depends on the smoothing parameter $\lambda$, and on the width of the control limits, which is determined by $c$ in formulas (4.6) and (4.7). These parameters need to be chosen with care. In the following, we will present some considerations on how to choose $\lambda$ and $c$. We will make use of the asymptotic LCL and UCL which appeared in formulas (4.6) and (4.7). We determine the value of the two parameters by fixing two points on the ARL curve. The first point that will be fixed is the desired in-control ARL. Fixing the in-control ARL is related to producer’s risk. If the process is in control, an out-of-control signal is unwanted. The producer will suffer a certain loss if it accidentally happens that the process is stopped as the result of a false out-of-control signal. This demand on the ARL curve is usually formulated in terms of a high minimal in-control ARL. The second point that is fixed on the ARL curve, relates to consumer’s risk. It is desirable for the customer to have a low ARL if the products are of unacceptable quality. Hence, the second point that will be fixed is usually formulated in terms of a maximal ARL when there is a certain large shift in the mean. By fixing this second point, the chart will be designed for detecting a shift in the mean of a certain size as fast as possible.

In Figure 4.1, iso-$L_{W_X}(0, 0)$ curves are drawn for various values of the in-control ARL. All combinations of $\lambda$ and $c$ on one curve yield the same
 CHAPTER 4. EWMA-TYPE CONTROL CHARTS

producer’s risk.

![Iso-LW_0_0 curves for various in-control ARLs](image)

**Figure 4.1:** Iso-\(L_{W_X}(0,0)\) curves for various in-control ARLs

From Figure 4.1, it is not clear which point on an iso-\(L_{W_X}(0,0)\) curve to choose. Therefore, for shifts in the mean ranging from 0 to 3, the point of the iso-\(L_{W_X}(0,0)\) curve is determined that yields the minimal out-of-control ARL, thereby minimizing the consumer’s risk. This was done for each of the six choices of iso in-control ARL curves. This results in six curves in the \(c, \lambda, \delta\) space. The projection of these curves on the \(\lambda, \delta\) plane is shown in Figure 4.2. The projections in the \(c, \delta\) plane is depicted in Figure 4.3.

With the aid of Figures 4.2 and 4.3, the EWMA chart can be designed to have minimal out-of-control ARL for a predetermined shift in the mean, given a certain in-control ARL.

In Figure 4.4, the ARL curve of an EWMA chart that is most sensitive to a shift of size \(1\sigma_X\) in the mean of \(E(X_t)\), given an in-control ARL of 370.4, is depicted. The parameters of \(c = 2.7878\) and \(\lambda = 0.1417\) were determined using Figures 4.2 and 4.3 as being the optimal choice of parameters for detecting a shift of \(1\sigma_X\) as soon as possible with the EWMA control chart. In addition, the ARL curve of the Shewhart chart for independent observations with the same in-control ARL is depicted. This curve was drawn earlier in Figure 3.1 of the previous chapter.
Figure 4.2: Optimal choice of $\lambda$ for each $\delta$.

Figure 4.3: Optimal choice of $c$ for each $\delta$. 
CHAPTER 4. EWMA-TYPE CONTROL CHARTS

Figure 4.4: ARL curves of the Shewhart chart and an EWMA chart for independent observations.

From Figure 4.4, we conclude that a shift of size $1\sigma_X$ in the mean of the observations is, on average, detected much earlier on the EWMA chart. The out-of-control ARL of the EWMA chart at $\delta = 1$ is 9.58, whereas the out-of-control ARL of the Shewhart chart for independent observations is 43.89. However, for values of $\delta > 2.6$, the Shewhart chart is slightly more efficient. The difference is so small that this is not visible in Figure 4.4.

In conclusion, the EWMA is more sensitive to small shifts in the mean of a series of observations, whereas the Shewhart chart is a little more sensitive in detecting large shifts in the mean. In Section 4.6, the sensitivity of the EWMA control chart to small changes in the mean is explained.

Lucas and Saccucci (1990) describe three enhancements of the EWMA chart that were presented earlier by Lucas and Crosier (1982) for use with the CUSUM chart. These include a so-called FIR feature, which makes the control chart more sensitive at startup, a combined Shewhart EWMA chart, which is sensitive in detecting both large and small shifts in $\mu$ and robust EWMA, that provides protection against outliers.

In the next section, we will investigate how the ARL of the EWMA control chart is affected by AR(1) dependence, when the chart is designed as if the data were independent.
4.2 Effect of ignoring serial correlation

In this section, we will simulate the effect of ignoring AR(1) dependence in the data on the behavior of the EWMA control chart. To this end, we will assume that an EWMA control chart has been designed for independent data. We will use the same values for the parameters $c$ and $\lambda$ as in the previous section. An SPC practitioner who uses such a chart (un)knowingly with AR(1) data, will interpret an out-of-control signal as in Figure 4.4. For example: an out-of-control signal will, on average, occur once every 370 observations if the process is actually in control, and an out-of-control signal will occur, on average, every 10 observations if there is a shift of one standard deviation in the mean. However, if the data is correlated, these interpretations are not valid anymore. This is illustrated by Table 4.1, where we tabulated for selected values of $\phi$ the in-control ARL and the out-of-control ARL when there has been a shift of size $1\sigma_Y$ in the mean of AR(1) data. The ARLs are simulated using 100,000 replications, except for the ARL(0) corresponding to $\phi = -0.3$; there we used 1,000 replications. The bracketed numbers are the corresponding standard errors.

Table 4.1: ARL values of an i.i.d. EWMA chart, applied to various AR(1) processes.

<table>
<thead>
<tr>
<th>$\phi$</th>
<th>$\sigma_Y$</th>
<th>$\text{E}[\bar{MR}/d_2(2)]$</th>
<th>ARL(0 $\sigma_Y$)</th>
<th>ARL(1 $\sigma_Y$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.9</td>
<td>2.2942</td>
<td>3.1623</td>
<td>3.1623</td>
<td>3.1623</td>
</tr>
<tr>
<td>-0.6</td>
<td>1.2500</td>
<td>1.5811</td>
<td>1.5811</td>
<td>1.5811</td>
</tr>
<tr>
<td>-0.3</td>
<td>1.0483</td>
<td>1.1952</td>
<td>33119.69</td>
<td>11.96</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.0</td>
<td>1.0000</td>
<td>1.0000</td>
<td>365.56</td>
<td>9.34</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.3</td>
<td>1.0483</td>
<td>0.8771</td>
<td>42.55</td>
<td>7.42</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.6</td>
<td>1.2500</td>
<td>0.7906</td>
<td>12.64</td>
<td>5.84</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.9</td>
<td>2.2942</td>
<td>0.7255</td>
<td>5.86</td>
<td>3.59</td>
</tr>
</tbody>
</table>

The entries in Table 4.1 have been arrived at assuming that the standard...
deviation of the observations is estimated using $\overline{MR}/d_2(2)$. Comparing the true value of $\sigma_Y$ (column 2) with the expectation of its estimator (column 3), we noted earlier in Section 3.3 that $\overline{MR}/d_2(2)$ is positively biased in case of negative $\phi$, and negatively biased in case of positive $\phi$. Hence, the control limits for the EWMA chart will be too wide in case of negative autocorrelation. In fact, the width of the control limits for $\phi = -0.9$ and $\phi = -0.6$ turned out to be so large that it was not feasible to simulate the ARL in these cases. In case of $\phi = -0.3$, simulating the run length was so time consuming that it was necessary to cut down the number of replications to 1,000 in the in-control situation.

Despite the missing results for $\phi = -0.9$ and $\phi = -0.6$, the warning from Table 4.1 is compelling. Serial correlation seems to have an even stronger effect on EWMA control charts than on Shewhart control charts. This agrees with the findings of Harris and Ross (1991). Negative autocorrelation makes the control chart very insensitive, positive autocorrelation will result in many false out-of-control signals. An out-of-control signal on an EWMA control chart that was designed for independent observations, but used with AR(1) data is thus seriously misinterpreted.

One important reason for the misinterpretation is the bias in the estimation of $\sigma_Y$, a problem that is relatively easily resolved. The question remains how an out-of-control signal on an EWMA control chart that accounts for serial correlation must be interpreted.

In the Sections 4.3 through 4.5, we will discuss EWMA control charts that account for AR(1) dependence in the data. In Section 4.3, it is discussed how to modify the limits of the EWMA chart to allow for first-order autoregressive correlation in the data. In Section 4.4, the EWMA residuals chart is discussed, and in Section 4.5 the EWMA control chart for modified residuals is discussed. In Section 4.6, the ARL behavior of these charts will be compared, assuming that all model parameters are known.

### 4.3 The modified EWMA chart

In the following sections, we return to the case in which successive observations are correlated. As in the previous chapter, the AR(1) case will be discussed. In Appendix A, ARL tables of control charts for other time series models are presented.

In Section 3.4, we discussed how the Shewhart chart can be modified to account for AR(1) data. Schmid (1997b) employs the same reasoning for the EWMA control chart. In this section, we will consider this approach.
4.3. **THE MODIFIED EWMA CHART**

Compared to the EWMA chart for i.i.d. observations, the limits of the EWMA chart for correlated observations are widened to account for the increase in the variance. A second adjustment is required due to the fact that the ARL behavior of an EWMA chart of independent observations is not the same as that of an EWMA chart for correlated observations. For a proper evaluation of the effect of serial correlation on the EWMA chart, the limits will be adjusted so that the in-control ARL of the EWMA chart for dependent observations equals a predefined in-control ARL of the EWMA chart for independent observations. The resulting control chart is called the *modified EWMA chart*.

The statistic that will be plotted in the control chart at time $t$ is an EWMA of the correlated data, and will be denoted by $W_{Y,t}$. In Zhang (1998) it was shown that if the sequence $\{Y_t\}$ is stationary, then $\{W_{Y,t}\}$ is asymptotically stationary. Successive realizations of $\{W_{Y,t}\}$ are generated by

$$W_{Y,t} = \lambda Y_t + (1 - \lambda)W_{Y,t-1},$$

where the sequence $\{Y_t\}$ consists of AR(1) observations, generated by model (2.3)

$$Y_t - \mu_t = \phi(Y_{t-1} - \mu_{t-1}) + \varepsilon_t \quad \text{for} \quad t \in \mathbb{Z},$$

where $\{\varepsilon_t\}$ is a sequence of i.i.d. disturbances, $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$ for $t \in \mathbb{Z}$.

Suppose that a special cause of variation may occur at some unknown time $T$, resulting in a persistent shift in $\text{E}(Y_t) = \mu_t$ of size $\delta \sigma_Y$:

$$\mu_t = \begin{cases} 
\mu & \text{for } t < T \\
\mu + \delta \sigma_Y & \text{for } t \geq T.
\end{cases} \quad (4.9)$$

We will investigate how quickly, on average, such a change is detected by monitoring the sequence $\{W_{Y,t}\}$.

For determining the control limits for the EWMA chart for AR(1) data,
we derive the variance of $W_{Y,t}$.

\[
\text{Var}(W_{Y,t}) = \text{Var} \left( \lambda \sum_{i=0}^{t-1} (1 - \lambda)^i Y_{t-i} + (1 - \lambda)^t W_{Y,0} \right) \\
= \frac{\sigma^2}{1 - \phi^2} \left( \frac{\lambda}{2 - \lambda} \right) \left[ 1 - (1 - \lambda)^{2t} \right] + \\
+ 2\lambda^2 \frac{\sigma^2}{1 - \phi^2} \sum_{i=0}^{t-2} \sum_{j=i+1}^{t-1} (1 - \lambda)^i \phi^j \phi^{-i-1} \\
= \frac{\sigma^2}{1 - \phi^2} \left( \frac{\lambda}{2 - \lambda} \right) \left[ 1 - (1 - \lambda)^{2t} \right] + \\
+ 2\lambda^2 \frac{\sigma^2}{1 - \phi^2} \sum_{i=0}^{t-2} \left( \frac{1 - \lambda}{\phi} \right)^i \sum_{j=i+1}^{t-1} ((1 - \lambda)\phi)^j \\
= \frac{\sigma^2}{1 - \phi^2} \left( \frac{\lambda}{2 - \lambda} \right) \left[ 1 - (1 - \lambda)^{2t} \right] + \\
+ 2 \frac{\sigma^2}{1 - \phi^2} \left( \frac{\lambda^2}{1 - \phi(1 - \lambda)} \right) \left\{ \phi(1 - \lambda) \sum_{i=0}^{t-2} (1 - \lambda)^{2i} + \right. \\
\left. - (\phi(1 - \lambda))^t \sum_{i=0}^{t-2} \left( \frac{1 - \lambda}{\phi} \right)^i \right\} \\
= \frac{\sigma^2}{1 - \phi^2} \left( \frac{\lambda}{2 - \lambda} \right) \left[ 1 - (1 - \lambda)^{2t} \right] + \\
+ 2 \frac{\sigma^2}{1 - \phi^2} \left( \frac{\lambda}{2 - \lambda} \right) \left( \frac{\phi(1 - \lambda)}{1 - \phi(1 - \lambda)} \right) \left[ 1 - (1 - \lambda)^{2t-2} \right] \\
- 2 \frac{\sigma^2}{1 - \phi^2} \left( \frac{\lambda^2}{1 - \phi(1 - \lambda)} \right) \left( \frac{\phi^{t+1}(1 - \lambda)^t}{\phi + \lambda - 1} \right) \left[ 1 - \left( \frac{1 - \lambda}{\phi} \right)^{t-1} \right]
4.3. **THE MODIFIED EWMA CHART**

\[
= \frac{\sigma^2_e}{1 - \phi^2} \left( \frac{\lambda}{2 - \lambda} \right) \left( \frac{1 + \phi(1 - \lambda)}{1 - \phi(1 - \lambda)} \right) \left[ 1 - (1 - \lambda)^2t \right] + \\
+ 2 \frac{\sigma^2_e}{1 - \phi^2} \left( \frac{\lambda}{2 - \lambda} \right) \left( \frac{\phi(1 - \lambda)}{1 - \phi(1 - \lambda)} \right) \left[ (1 - \lambda)^{2t} - (1 - \lambda)^{2t-2} \right] \\
- 2 \frac{\sigma^2_e}{1 - \phi^2} \left( \frac{\lambda^2}{1 - \phi(1 - \lambda)} \right) \left( \frac{\phi^2(1 - \lambda)^t}{\phi + \lambda - 1} \right) \left( \phi^{t-1} - (1 - \lambda)^{t-1} \right).
\]

The variance of \( W_{Y,t} \) is not constant over time. However,

\[
\text{Var}(W_{Y,t}) \approx \frac{\sigma^2_e}{1 - \phi^2} \left( \frac{\lambda}{2 - \lambda} \right) \left( \frac{1 + \phi(1 - \lambda)}{1 - \phi(1 - \lambda)} \right) \quad \text{for large } t.
\]

Control limits for monitoring the sequence \( \{W_{Y,t}\} \) are of the form

\[
\begin{align*}
\text{LCL}_t &= \mu - c\sigma_{W_{Y,t}} \\
\text{UCL}_t &= \mu + c\sigma_{W_{Y,t}},
\end{align*}
\]

where either the exact standard deviation of \( \{W_{Y,t}\} \) can be used, which will result in control limits that vary over time, or the asymptotic standard deviation, which results in constant control limits. Surprisingly, simulation studies performed by Schmid (1997b) indicate that utilizing the exact variance for the control limits does not always lead to better ARL behavior. His results also show that there is essentially no difference in the ARL behavior. Therefore, in the remainder of this section, we will work with the asymptotic variance.

In order to implement the EWMA chart the two parameters \( \lambda \) and \( c \) have to be chosen. Schmid (1997b) presents combinations for \( \lambda \) and \( c \) that yield an in-control ARL of 500 for various values of \( \phi \). For other in-control ARLs, combinations of \( c \) and \( \lambda \) can be determined by fixing two points on the ARL curve of the control chart. For the ARL of the modified EWMA chart in case of AR(1) observations we present the following considerations, which are inspired by an excellent article by VanBrackle and Reynolds (1997), where the ARL curve of the modified EWMA chart for ARMA(1,1) data was presented. In Schmid and Schöne (1997) some theoretical results on the EWMA control chart in the presence of autocorrelation are presented.
Let $L_{W_Y}(\delta, u, v)$ denote the ARL of the modified EWMA chart if the first observation is taken after the mean has shifted by an amount of $\delta \sigma_Y$, and that the value of EWMA statistic is $u$ at the first observation, while the corresponding AR(1) observation is $v$. Then the following holds

$$L_{W_Y}(\delta, u, v) =$$

$$= 1 + \int_{\{y:y+(1-\lambda)u-\mu| \leq \sigma_{W_Y}\}} L_{W_Y}(\delta, (1-\lambda)u + \lambda y, y) \times$$

$$\times f(y - \phi v - (1 - \phi)(\mu + \delta \sigma_Y)) \, dy$$

$$= 1 + \frac{1}{\lambda} \int_{\mu+\sigma_{W_Y}}^{\mu-\sigma_{W_Y}} L_{W_Y}(\delta, w, \frac{w - (1 - \lambda)u}{\lambda}) \times$$

$$\times f\left(w - (1 - \lambda)u + \lambda y - \phi v - (1 - \phi)(\mu + \delta \sigma_Y)\right) \, dw,$$

where

$$w = (1 - \lambda)u + \lambda y,$$

and $f(\cdot)$ is the probability density function of the disturbances. Compared to the derivation of the ARL curve of the EWMA chart for independent observations in (4.8), the expression above is more complicated, since the information that is needed to go from one time point to another is no longer one-dimensional. In addition to the value of the EWMA statistic, it is also necessary to know the value of the AR(1) observation.

If the first observation is taken at the time of the shift, the corresponding ARL function is obtained as follows

$$L_{W_Y}(\delta, u^*, v^*) = 1 + \frac{1}{\lambda} \int_{\mu+\sigma_{W_Y}}^{\mu-\sigma_{W_Y}} L_{W_Y}(\delta, w^*, y^*) \times$$

$$\times f(y^* - \phi v^* - (1 - \phi)(\mu - \delta \sigma_Y)) \, dw^*, $$

where $u^*$ is the last observed value of the EWMA statistic before the shift in the mean occurred, and $v^*$ is the corresponding AR(1) observation. Analogously, $w^*$ is the value of the EWMA statistic at the time of the shift. The corresponding AR(1) observation is denoted by $y^*$ and equals
4.4. THE EWMA CHART OF RESIDUALS

\[ y^* = \frac{w^* - (1 - \lambda)u^*}{\lambda}. \]

The dependence of \( L_{W_Y}^*(\delta, u^*, v^*) \) on \( u^* \) and \( v^* \) can be 'averaged out' to obtain

\[ L_{W_Y}^*(\delta) = \int_{\mu - \sigma_{W_Y}}^{\mu + \sigma_{W_Y}} \int_{-\infty}^{\infty} L_{W_Y}^*(\delta, u^*, v^*) \ k(v^*) g(u^*) \ dv^* du^*, \]

where \( k(\cdot) \) is the probability density function of the last AR(1) observation just before the occurrence of the shift, and \( g(\cdot) \) is the truncated probability density function of the last observation of the EWMA just before the shift.

The presentation and the discussion of the ARL curves of the modified EWMA chart are deferred to Section 4.6. In that section, the ARL curves of the modified EWMA chart are compared to the ARL curves of the EWMA chart of residuals and the EWMA chart of modified residuals for various values of the AR-parameter \( \phi \). But first we discuss the EWMA chart of residuals in Section 4.4, and the EWMA chart of modified residuals in Section 4.5.

4.4 The EWMA chart of residuals

In this section, the EWMA chart for the mean of residuals of a fitted AR(1) process is discussed. As in Section 3.5, residuals \( \{e_t\} \) are computed as follows, assuming that the mean of the AR(1) observations equals \( \mu \):

\[ e_t = Y_t - \mu - \phi(Y_{t-1} - \mu). \]

If \( \mu \) and \( \phi \) are not known, they have to be replaced by appropriate estimates. However, we will assume that \( \mu \) and \( \phi \) are known. Concerning the effect of parameter estimation on the performance of control charts, as yet little work has been done. This is also outside the scope of this thesis.

If \( \mu \) and \( \phi \) are known, \( e_t = E(e_t) + e_t \). If \( E(Y_t) \) shifts from \( \mu \) to \( \mu + \delta \sigma_Y \) at time \( T \), we have for the expectation of the residuals

\[
E(e_t) = \begin{cases} 
0 & \text{for } t < T \\
\delta \sigma_Y & \text{for } t = T \\
(1 - \phi)\delta \sigma_Y & \text{for } t > T,
\end{cases}
\]
CHAPTER 4. EWMA-TYPE CONTROL CHARTS

where we recall from Chapter 3 that

\[ \sigma_Y = \frac{\sigma_\varepsilon}{\sqrt{1 - \phi^2}} \]

when \( \{Y_t\} \) is generated by an AR(1) model with \('white noise\'\) disturbances \( \{\varepsilon_t\} \), with \( \varepsilon_t \sim i.i.d. \mathcal{N}(0, \sigma_\varepsilon^2) \). The EWMA of the residuals at time \( t \) is denoted by \( W_{e,t} \) and is computed as follows

\[ W_{e,t} = \lambda \varepsilon_t + (1 - \lambda)W_{e,t-1}. \]

Assuming that \( \mu \) and \( \phi \) are known, the variance of \( W_{e,t} \) is, analogously to Equation (4.3), derived as

\[ \sigma^2_{W_{e,t}} = \sigma^2_\varepsilon \left( \frac{\lambda}{2 - \lambda} \right) \left[ 1 - (1 - \lambda)^{2t} \right]. \]

As an asymptotic approximation,

\[ \sigma^2_{W_{e,t}} \approx \sigma^2_\varepsilon \left( \frac{\lambda}{2 - \lambda} \right) \]

may be used. Control limits for the EWMA residuals chart are of the form

\[ \begin{align*}
LCL_t &= \mu - c\sigma_{W_{e,t}} \\
UCL_t &= \mu + c\sigma_{W_{e,t}},
\end{align*} \]

where either the exact or the asymptotic expression for the standard deviation of \( W_{e,t} \) can be used. The two design parameters \( \lambda \) and \( c \) are again chosen by fixing two points on the ARL curve.

In Chapter 3, it was argued that monitoring a serially correlated process by monitoring the residuals of a fitted time series model is appealing. If the time series model is appropriate for the data, then the residuals will be approximately uncorrelated.

In Section 4.1, the ARL curve of a sequence of independent observations was derived. Since under the assumptions made above, the sequence \( \{e_t\} \) consists of independently distributed random variables, the ARL curve of the EWMA residuals chart is derived analogously. That is, the ARL of the EWMA residuals chart that started in \( u \), which is denoted by \( L_{W_{e}}(\delta, u) \), satisfies the following integral equation if the first observation is taken after the shift in the mean has occurred.
4.5. THE EWMA CHART OF MODIFIED RESIDUALS

\[ L_{We}(\delta, u) = 1 + \frac{1}{\lambda} \int_{-c\sigma_{We}}^{c\sigma_{We}} L_{We}(\delta, v) f\left(\frac{v - (1 - \lambda)u}{\lambda} - (1 - \phi)\sigma_Y\right) dv, \]

where \( v \) is the value of the EWMA statistic that succeeds \( u \). If the first observation is taken at the time of the shift, \( L_{We}^*(\delta, u^*) \), the ARL of the EWMA residuals chart that started in \( u^* \), satisfies

\[ L_{We}^*(\delta, u^*) = 1 + \frac{1}{\lambda} \int_{-c\sigma_{We}}^{c\sigma_{We}} L_{We}(\delta, v^*) f\left(\frac{v^* - (1 - \lambda)u^*}{\lambda} - \delta\sigma_Y\right) dv^*, \]

where \( v^* \) is the value of the EWMA of residuals at the time of the shift. Note that \( u^* \) is the last value of the EWMA statistic just before the shift. The dependence of \( L_{We}^*(u^*) \) on \( u^* \) can be ‘integrated out’ to obtain

\[ L_{We}^{**}(\delta) = \int_{-c\sigma_{We}}^{c\sigma_{We}} L_{We}^*(\delta, u^*) g(u^*) du^*, \]

where \( g(\cdot) \) is the density function of the EWMA statistic that is observed just before the shift. The resulting ARL curves are drawn, together with the ARL curves of the modified EWMA chart and the EWMA modified residuals chart in Section 4.6. In the next section, the EWMA modified residuals chart is discussed.

4.5 The EWMA chart of modified residuals

In Chapter 3, a modified residual was proposed that can be used to monitor serially correlated data. In this section, we discuss monitoring the mean of an AR(1) process with an EWMA control chart of the modified residuals.

Recall from Section 3.6 that the modified residuals \( \{u_t\} \) of an AR(1) process are computed as

\[ u_t \equiv Y_t - \phi Y_{t-1} + \hat{\mu}_t, \quad (4.10) \]

where \( \hat{\mu}_t \) is an estimator of the level of the process that quickly responds to changes in the ‘local mean’ \( \hat{\mu}_t \). Simulation studies (which are not included in this thesis) have shown that an EWMA of the observations is a better choice than an unweighted moving average, in terms of ARL behavior of the resulting control chart.

Assuming that \( \mu \) and \( \phi \) are known, and that \( \hat{\mu}_t \) is an EWMA of previous observations with parameter \( \lambda' \), the expectation of \( u_t \) equals
CHAPTER 4. EWMA-TYPE CONTROL CHARTS

\[
E(u_t) = \begin{cases} 
(1 - \phi)\mu + \phi \mu \sum_{i=0}^{\infty} (1 - \lambda)^i = \mu & \text{for } t < T \\
\mu + (1 + \phi \lambda)\delta Y & \text{for } t = T \\
\mu + (1 - \phi)\delta Y + \phi \delta Y \sum_{i=0}^{t-T} (1 - \lambda)^i & \text{for } t > T,
\end{cases}
\]

where \( T \) is again the unknown time point when a shift in \( E(Y_t) \) occurs. Note that the mean of the sequence \( \{u_t\} \) is \( \mu \) before a possible shift in the mean of the process. After a shift in the mean of the process, the mean of \( \{u_t\} \) converges to \( \mu + \delta Y \) for \( t \gg T \).

The EWMA of \( \{u_t\} \) at time \( t \) is denoted by \( W_{u,t} \) and is computed as

\[
W_{u,t} = \lambda u_t + (1 - \lambda)W_{u,t-1} \quad \text{for } t = 0, 1, 2, \ldots.
\]

The ARL curves for the EWMA chart of modified residuals of AR(1) data are obtained by simulation. In Section 4.6, these are compared to the ARL curves of the modified EWMA chart and the EWMA chart of residuals for various values of \( \phi \).

4.6 Discussion

In the previous sections, various EWMA control charts for monitoring the mean of AR(1) observations were discussed. In the next subsection, the ARL curves corresponding to these control charts are presented. We will see that the ARL behavior of EWMA-type control charts is in general much better than the ARL behavior of Shewhart-type control charts for small shifts in the mean. This was also observed in Section 4.1 of this chapter. In Subsection 4.6.2, we will explain why this is the case.

4.6.1 ARL comparison

In the previous three sections, respectively the modified EWMA, the EWMA chart of residuals and the EWMA chart of modified residuals were discussed. For each of the charts, considerations on the computation of ARL curves were presented. However, compared to the ARL curves in Chapter 3, the computation time needed to compute the ARL curves of the EWMA-type control charts rose dramatically. Therefore, we decided to
evaluate the ARL curves of the EWMA charts by means of simulation. In Figures 4.5 through 4.10, the ARL curves of the three charts are compared to the ARL curve of the EWMA control chart for the mean of independent observations, for values of $\phi = -0.9, -0.6, -0.3, 0.3, 0.6, 0.9$. The value of the EWMA parameter $\lambda$ was chosen equal to 0.2. For the EWMA chart of modified residuals, we took $\lambda' = 0.1$, as in Section 3.6. Each of the curves in Figures 4.5 through 4.10 consists of 101 points, and each point is the mean of 100,000 run lengths.

From Figures 4.5 through 4.10, we conclude that, generally speaking, the effect of first-order autocorrelation on EWMA charts is the same as on Shewhart-type charts. Compared to the i.i.d case, the ARL behavior is better for negative $\phi$, whereas ARL behavior is worse for positive $\phi$.

The difference in ARL behavior of the three charts is small for negative $\phi$. Also, for positive $\phi$, the ARL behavior of the modified EWMA chart and the ARL behavior of the EWMA of modified residuals does not differ much. For large $\phi$, the EWMA chart of residuals is performing worse than the other two. These conclusions agree with the findings of Schmid (1997a) and Schmid (1997b).
**Figure 4.6:** Various ARL curves for AR(1) process with \( \phi = -0.6 \), compared to the ARL curve for the i.i.d. case (\( \phi = 0 \)).

**Figure 4.7:** Various ARL curves for AR(1) process with \( \phi = -0.3 \), compared to the ARL curve for the i.i.d. case (\( \phi = 0 \)).
4.6. DISCUSSION

Figure 4.8: Various ARL curves for AR(1) process with $\phi = 0.3$, compared to the ARL curve for the i.i.d. case ($\phi = 0$).

Figure 4.9: Various ARL curves for AR(1) process with $\phi = 0.6$, compared to the ARL curve for the i.i.d. case ($\phi = 0$).
CHAPTER 4. EWMA-TYPE CONTROL CHARTS

![Graph showing ARL curves for various control charts.](image)

**Figure 4.10:** Various ARL curves for AR(1) process with \( \phi = 0.9 \), compared to the ARL curve for the i.i.d. case (\( \phi = 0 \)).

Note that for small \( \phi \) the ARL behavior of EWMA-type control charts appears to be better than the ARL behavior of Shewhart-type charts (compare to Figures 3.21 through 3.26 of the previous chapter). However, for large \( \phi \), the ARL performance of EWMA-type control charts is comparable to Shewhart-type control charts.

Hence, for small to moderate levels of first-order autoregressive serial correlation, it is advisable to use an EWMA-type control chart to monitor the process for changes in the mean. However, for larger values of \( \phi \), a Shewhart-type control chart might be preferred because it is easier to compute and to interpret.

### 4.6.2 Relationship between the EWMA and the modified Shewhart chart

In this subsection the modified Shewhart chart for AR(1) data is compared to the EWMA control chart for independent observations. It turns out that they are very similar. In Figure 4.4, it is shown that the ARL behavior of the EWMA control chart is better than the ARL performance of the Shewhart chart for independent observations. Therefore, one might expect
4.6. DISCUSSION

that the ARL performance of the modified Shewhart chart for AR(1) data is also better than that of the Shewhart chart for independent observations. However, in Chapter 3, it was argued that the ARL performance of the modified Shewhart chart for AR(1) data is comparable to that of the Shewhart chart for independent observations. In this subsection, it is explained why the ARL performance of the modified Shewhart chart for AR(1) data is not as good as the ARL performance of the EWMA chart for independent observations.

Suppose that we have a sequence of i.i.d. observations \( \{X_t\} \), which satisfy (3.1) and (3.2). The EWMA statistic at time \( t \) is denoted by \( W_{X,t} \) and is constructed as follows

\[
W_{X,t} = (1 - \lambda)W_{X,t-1} + \lambda X_t \quad \text{for } t = 1, 2, \ldots, \tag{4.11}
\]

where \( \lambda \in (0, 1) \). The EWMA chart may be started by setting \( W_{X,0} \) equal to a target value or (an estimator of) \( \mu \). If \( W_{X,0} = \mu \), it is easy to verify that \( \mathbb{E}(W_{X,t}) = \mu \) for \( t = 0, 1, 2, \ldots, T \). For \( t \geq T \) we have

\[
\mathbb{E}(W_{X,t}) = \mu + \left\{ 1 - (1 - \lambda)^{t-T+1} \right\} \delta \sigma_X,
\]

which approximately equals \( \mu + \delta \sigma_X \) for \( t \gg T \). Hence, \( \{\mathbb{E}(W_{X,t})\} \), the sequence of expected values of the EWMA statistic, approximately mimics \( \{\mathbb{E}(X_t)\} \).

That there is a relation between (4.11) and the AR(1) model (2.3) becomes clear if we subtract \( \mu_t \) on both sides of Equation (4.11):

\[
W_{X,t} - \mu_t = (1 - \lambda)(W_{X,t-1} - \mu_t) + \lambda(X_t - \mu_t). \tag{4.12}
\]

Note that \( \{\lambda(X_t - \mu_t)\} \) may be considered a white noise process with variance \( \lambda^2 \sigma_X^2 \). As long as \( \mu_t = \mu \) for \( t = 0, 1, 2, \ldots, \), we conclude that computing the EWMA statistic is equivalent to converting a sequence of i.i.d. observations into an AR(1) sequence with AR parameter \( \phi = 1 - \lambda \). Moreover, using an EWMA control chart is in a certain sense equivalent to monitoring AR(1) observations using modified Shewhart chart. In Lucas and Saccucci (1990) it was shown that the properties of the EWMA chart are very close to those of CUSUM schemes. Small shifts in the mean are, on average, more quickly detected on an EWMA control chart than on a standard Shewhart control scheme.

The argument above seems somewhat in contradiction with the conclusions of Section 3.4. There we observed that for \( \phi = 0.9 \) the modified
Shewhart control chart for AR(1) data is not very sensitive in detecting small changes in the mean. A value of $\phi = 0.9$ corresponds to a value of $\lambda = 0.1$. For this value of $\lambda$, an EWMA chart is much more sensitive than a Shewhart control chart for detecting small shifts in the mean of a sequence of independent observations. The discrepancy between these results can be explained by studying a signal-to-noise ratio: a number that relates the size of the shift to the standard deviation of the process. This ratio allows us to compare shifts in means in processes with different variances. In Section 3.5.2, the signal-to-noise ratio was defined as the size of the shift divided by the standard deviation of the process.

For the modified Shewhart chart, the size of the shift is $\delta \sigma_Y$ after time $T$. The standard deviation of the AR(1) process is $\sigma_Y$. Hence, the signal-to-noise ratio is $\delta$.

In the case of the EWMA control chart, the size of the shift in $\text{E}(W_{X,t})$ approximately equals $\delta \sigma_X$ for $t \gg T$. The variance of the EWMA statistic is

$$\text{Var}(W_{X,t}) = \sigma_X^2 \left( \frac{\lambda}{2 - \lambda} \right) \left\{ 1 - (1 - \lambda)^{2t} \right\}.$$ 

Hence, for large $t \gg T$, the signal-to-noise ratio equals approximately $\delta \sqrt{(2 - \lambda)/\lambda}$, which is larger than $\delta$ for $0 < \lambda < 1$. A popular choice of $\lambda$ for the EWMA chart is $\lambda = 0.1$. In this case, the signal-to-noise ratio is approximately $\delta \sqrt{19}$ for large $t \gg T$.

Hence, computing the EWMA of a sequence of i.i.d. observations leaves the pattern of expectations approximately unaltered, but improves the signal-to-noise ratio. This is combined with the introduction of first-order positive autocorrelation. From the ARL curves in the previous subsection, we learned that first-order positive autocorrelation has a (small) negative effect on the ARL performance. Apparently, this effect is offset by the positive effect of the improved signal-to-noise ratio. We conclude that the efficiency of the EWMA control chart is not the result of the autocorrelation that is introduced. It is the improvement of the signal-to-noise ratio that makes the EWMA control chart efficient.