



Research Report
Department of Statistics
Göteborg University
Sweden

**Optimal surveillance
Based on exponentially
weighted moving averages**

**Marianne Frisé
Christian Sonesson**

**Research Report 2002:1
ISSN 0349-8034**

Mailing address:	Fax	Phone	Home Page:
Dept of Statistics	Nat: 031-773 12 74	Nat: 031-773 10 00	http://www.stat.gu.se/stat
P.O. Box 660	Int: +46 31 773 12 74	Int: +46 31 773 10 00	
SE 405 30 Göteborg			
Sweden			

Optimal Surveillance

Based on

Exponentially Weighted Moving Averages

Marianne Frisé and Christian Sonesson

Statistical surveillance is used to detect an important change in a process as soon as possible after it has occurred, with control of false alarms. The EWMA, exponentially weighted moving average, method for surveillance is used in different areas, such as industry, economy and medicine. Three optimality criteria of surveillance are studied and the implications are described for the EWMA method and for suggested modifications.

The first criterion concerns the average run length to alarm, ARL. This is the most commonly used criterion. Results on ARL optimality for EWMA are demonstrated. Equal weight for old and recent observations give good ARL-properties but bad properties otherwise. Thus, uncritical use of this criterion should be avoided.

The second criterion is the ED criterion based on the minimal expected delay from change to detection. The full likelihood ratio method is optimal according to this criterion. Various approximations of this method turn out to be modifications of the EWMA method. Two of these modifications keep the EWMA statistic unchanged and just alter the alarm limits slightly. The approximations lead to a formula for the value of the weight parameter of the EWMA statistic. The usefulness of this formula is demonstrated. The conventional EWMA and the modifications are compared to the optimal full likelihood ratio method. No modification of EWMA is necessary for detection of large changes (where also the Shewhart

method is useful) but all the modifications give considerable improvement for small changes.

The third criterion is based on the minimax of the expected delay after a change with respect to the time of the change. It is demonstrated that the value of the smoothing parameter, which is optimal according to this criterion, agrees well with that of the ED criterion but not with that of the ARL criterion.

A restriction on the false alarm property is necessary. For the ARL criterion it is the ARL without any change. For the other two criteria we here use the false alarm probability. It is demonstrated that these two restrictions favor different methods.

KEY WORDS: Monitoring; Quality control; Stopping rule; Optimal; Minimax; Expected delay.

Marianne Frisé is Professor, Department of Statistics, Göteborg University, SE40530 Göteborg, Sweden. (E-mail: MarianneFrisen@Statistics.GU.SE) Christian Sonesson is graduate student at the same department. This work was supported by the Swedish Council for Research in the Humanities and Social Sciences, Grant F0473/2000. The authors thank Hans van Houwelingen for constructive discussions.

1. INTRODUCTION

Continual surveillance of time series, with the goal of detecting an important change in the underlying process as soon as possible after it has occurred is needed in many areas. An example is the surveillance of the foetal heart during labor described by Frisé (1992). An abnormality, caused by e.g. a lack of oxygen due to the umbilical cord around the neck of the foetus might happen at any time. If an alarm is given soon after the event has occurred it is possible to rescue the baby by e.g. a Cesarean section. Other applications in medicine are described in e.g. the special issue (no. 3, 1989) of "Statistics in Medicine". A review of methods for the surveillance in public health is given by Sonesson and Bock (2003). Environmetric control is described by e.g. Pettersson (1998). Applications in economics, and especially the surveillance of business cycles, are treated in e.g. the special issue (no. 3/4, 1993) of "Journal of Forecasting" and by Andersson (2002).

Several methods for surveillance have been suggested in the literature. Broad surveys and bibliographies on surveillance are given by Lai (1995), who concentrates on minimax properties of stopping rules and by Woodall and Montgomery (1999), who concentrate on control charts. Here we concentrate on the method for surveillance based on exponentially weighted moving averages, usually called EWMA. It is useful for the analysis of several important issues of inference in connection with surveillance since it has a simple form, can be expected to have good properties, and lacks the discontinuity of the CUSUM method. The EWMA method was introduced in the quality control literature by Roberts (1959) and the method has got much attention. This may be due to papers such as Robinson and Ho (1978), Crowder (1987), Ng and Case (1989), Lucas and Saccucci (1990) and Domangue and Patch (1991) in which positive reports on the quality of the method are given. Also, Srivastava and Wu (1997) pointed out the fact that the EWMA control chart can easily be used to estimate the current mean and to make optimal forecasts.

The concept of optimality in on-line surveillance is discussed by, e.g. Frisé and de Maré (1991) and Lai (1995). In a surveillance situation, there is no fixed data set and not even a fixed hypothesis to be tested. The choice of an optimality criterion is an interesting and important issue. Optimal EWMA is discussed by e.g. Crowder (1987), Lucas and Saccucci (1990) and Srivastava and Wu (1997). These papers all consider an optimality criterion based on the average run length when a change either happens immediately or not at all. Earlier studies on optimal EWMA have mainly focused on this criterion. Here, also other optimality criteria are applied to the EWMA method. The optimality condition based on minimal expected delay from change to detection, suggested by Girshick and Rubin (1952) and Shiryaev (1963) is used. The full likelihood ratio method is optimal according to this criterion. Various approximations of this method turn out to be modifications of the EWMA method. Also, a minimax criterion and its relation to the other two criteria are studied. The criteria are described in Section 4 and analyzed in the following sections.

Most studies on EWMA are for the two-sided case. Here we discuss the differences between one- and two-sided EWMA but concentrate on the one-sided case, where the inference logic is less complicated. Only surveillance of a process of independent and univariate observations is examined in order to focus on some critical issues of optimality.

In Section 2 some notations are given and the case studied is specified. In Section 3 some properties of some variants of the EWMA method are described. In Section 4 the criteria of optimality which are studied are described and analyzed. In Section 5 different constructions of optimal EWMA are described and motivated. In this section a large scale simulation study on different properties for EWMA and competitors are reported. Section 6 contains a discussion. The reliability of the simulation study is reported in the Appendix.

2. NOTATIONS AND SPECIFICATIONS

At each decision time $s, s=1, 2, \dots$ we want to discriminate between the two events $C(s)$, which requires an action, and $D(s)$, which does not. The decision at time s is based on $X_s = \{X(t): t = 1, 2, \dots, s\}$, where $X(t)$ is the observation made at time t . The observation may be an average or some other derived statistic. The random process that determines the state of the system is denoted by $\mu = \{\mu(t), t = 1, 2, \dots\}$. Given μ , the observations are assumed independent with the same known standard deviation, σ . At some unknown time point τ , there is a change in the distribution of $X(t)$. As in most literature, the case of a shift in the mean of a Gaussian random variable from a value μ^0 to another value μ^1 is considered. Only one-sided alternatives are considered here, where $\mu^1 > \mu^0$. Here, μ^0 and μ^1 are regarded as known values and without loss of generality we set $\mu^0 = 0$, the standard deviation $\sigma = 1$ and μ^1 is denoted μ . The process suddenly shifts at time τ and remains at the new level. That is,

$$\mu(t) = 0 \text{ for } t = 1, \dots, \tau-1 \text{ and } \mu(t) = \mu \text{ for } t = \tau, \tau+1, \dots$$

The time point, τ , when the change in the distributions of $X(t)$ occurs, is regarded as a random variable. We study the case of a constant intensity $v = P(\tau=t | \tau \geq t)$.

The distribution of τ is thus geometric with $P(\tau=t) = v(1-v)^{t-1}$. We aim to discriminate between the two events $C(s) = \{\tau \leq s\}$ and $D(s) = \{\tau > s\}$ by the set of observations X_s . We will consider different ways of constructing alarm sets $A(s)$ with the property that, when X_s is a subset of $A(s)$, there is an indication that the event $C(s)$ has occurred. The time of the first alarm is $t_A = \min\{s: X_s \in A(s)\}$. For the methods studied in this report this can be expressed as

$$t_A = \min\{s: p(X_s) > L(s)\},$$

where p is an alarm statistic and L is an alarm limit.

3. THE EWMA METHOD

The EWMA method for surveillance has an alarm statistic based on exponentially weighted moving averages,

$$Z_s = (1-\lambda)Z_{s-1} + \lambda X(s), \quad s=1, 2, \dots$$

where $0 < \lambda < 1$ and here, as usual, Z_0 is the target value μ^0 , which is set to zero. The alarm statistic is sometimes referred to as a geometric moving average. It can equivalently be written as

$$Z_s = \lambda(1-\lambda)^s \sum_{t=1}^s (1-\lambda)^{-t} X(t).$$

The EWMA statistic gives the most recent observation the greatest weight, and gives all previous observations geometrically decreasing weights. If λ is near zero, all observations have approximately the same weight. If λ is equal to one, only the last observation is considered as in the Shewhart method. Shewhart (1931) suggested that an alarm should be given at

$$t_A = \min\{s: X(s) > L\},$$

where L is a constant.

If no change ever occurs, the expected value of the EWMA statistic is $E(Z_s | \mu(t) \equiv 0) = 0$. If the change occurs immediately, we have $E(Z_s | \mu(t) \equiv \mu) = \mu(1-(1-\lambda)^s) \rightarrow \mu$ as $s \rightarrow \infty$. The variance of the statistic is $\sigma_{Z_s}^2 = \lambda[1-(1-\lambda)^{2s}]/(2-\lambda)$ which tends to $\sigma_Z^2 = \lambda/(2-\lambda)$ when $s \rightarrow \infty$. The rate of convergence of the variance depends on λ . For small values of λ it is very slow. However, the use of the asymptotic variance is very common and it can in fact be preferred to the exact version in some aspects. This procedure will be referred to as the ‘‘asymptotic version’’. When it is necessary to distinguish between the exact and the asymptotic version, the notations

EWMAe and EWMAa will be used. Most of the results below concern EWMAa. When not otherwise indicated it is this version which is analyzed. EWMAa will give an alarm at time t_A , where

$$t_A = \min\{s: Z_s > L\sigma_Z\},$$

where L is a constant. For EWMAe the exact standard deviation is used instead of the asymptotic. For the comparison with other methods in the following sections it is useful to express the alarm statistic with normalized weights $w_{EWMA}(s,t) = \lambda(1-\lambda)^{s-t}/[1-(1-\lambda)^s]$, which sum to one. The alarm condition is then

$$Z_{W_s} = \sum_{t=1}^s w_{EWMA}(s,t)X(t) > L_{EWMA}(s)$$

For EWMAa we have

$$L_{EWMAa}(s) = L\sigma_Z / [1-(1-\lambda)^s].$$

Small values of λ will be of special concern in Section 5.1 on ARL optimality. In the other end we have the large values of λ . For $\lambda=1$ the EWMA method is identical with the Shewhart method.

Since an alarm for EWMAe (obvious adjustment for EWMAa) is given if the statistic Z_s exceeds the alarm limit, $L\sigma_{Z_s}$, where L is a constant, we have that $P(Z_s > L\sigma_{Z_s}) = \Phi(-\{L-\mu/\sigma_{Z_s}\})$, where Φ is the normal distribution function. For this exact version this probability is constant. For $\lambda=1$ successive values of Z_s are independent. We have in that case $ARL^0 = 1/\Phi(-L)$ and $ARL^1 = 1/\Phi(-\{L-\mu\})$. For other values of λ , there is a dependency between successive values of Z_s , and the dependency increases when λ decreases. The value of L , which gives a desired ARL^0 , is a function of this correlation which in turn is a function of λ . For a fixed value of L the smallest value of ARL^0 is obtained for $\lambda=1$ and the largest when it approaches zero. Equivalently, for a fixed value of ARL^0 the largest value of L is obtained

when $\lambda=1$ and the smallest when it approaches zero.

The use of EWMA for autocorrelated processes is discussed by e.g. Lu and Reynolds Jr (1999), who compare application of the EWMA to the original observations with application to the residuals from an autoregressive model. Multivariate EWMA methods are discussed by e.g. Domangue and Patch (1991), Tsui and Woodall (1993), Lowry and Montgomery (1995), Gan (1995), Prabhu and Runger (1997) and Bodden and Rigdon (1999). Here, only surveillance of a process of independent and univariate observations is discussed in order to focus on some critical issues of optimality.

Two-sided EWMA methods can be achieved by running two one-sided surveillance procedures parallel and to make an alarm as soon as one of them signals a change. When the one-sided versions are symmetrical, this is equivalent to making an alarm as soon as $|Z_s| > L\sigma_z$. The investigations by Lucas and Saccucci (1990), Crowder (1989) and Srivastava and Wu (1997) use two-sided limits symmetrical around zero. The properties for one- and two-sided EWMA are not easily related because of different relations between successive decisions.

Champ, Woodall and Mohsen (1991), Gan (1995) and Gan (1998) use barriers for the alarm statistic. These authors use $\max\{0, Z_s\}$, respectively $\min\{0, Z_s\}$, in the one-sided surveillance in order to avoid bad properties in the “worst possible” case when the change occurs when the earlier observations least favour the detection of the change. This technique also has the advantage of the same simple relation between the one- and two-sided versions as for the CUSUM method. A formula by Kemp (1961) can be used for derivation of the ARL properties.

In this paper we will only deal with one-sided procedures in order to draw attention to certain optimality issues since optimality is most clearly expressed in this situation. However, it should be noticed that the results are not immediately translated to the two-sided case.

4. OPTIMALITY CRITERIA

The performance of a method for surveillance depends on the time point τ of the change. Sometimes it is appropriate to express the performance as a function of τ , as by Frisé (1992) and Frisé and Wessman (1999). Sometimes, however, a single criterion of optimality is needed. We will here study three such criteria, the first one often used in the quality control literature and in connection with EWMA and the two others often used in literature on general surveillance but seldom in the construction of optimal EWMA.

First, a measure which is often used in quality control, and which was suggested by Page (1954), is the average of the run length until the first alarm. The average run length until an alarm, when there is no change in the process under surveillance, is denoted by $ARL^0 = E(t_A | \mu(s) = 0)$. The average run length until detection of a true change (that occurred at the same time as the surveillance started) is denoted by $ARL^1 = E(t_A | \mu(s) = \mu) = E(t_A | \tau = 1)$. In the literature on quality control, optimality is often stated as minimal ARL^1 for fixed ARL^0 . This criterion is, for short, named the ARL criterion. Margavio, Conerly, Woodall and Drake (1995), Woodall and Montgomery (1999) and Carlyle, Montgomery and Runger (2000) stated that the use of the ARL criterion usually is recommended in spite of the known fact that the distributions of the alarm time are skewed. Margavio, et al. (1995) suggested that the whole distribution should be used. A time dependent limit can be chosen to give the desired distribution. By this, special properties such as fast initial alarms could be designed. However, the the possible problem of an influence of τ remains. For some situations, the expected delay of an alarm, given that there was no alarm earlier, is only slightly dependent on τ but this is not generally true. This will be demonstrated in Section 5.2. Degenerated methods, which would never be used in practice, give minimal ARL^1 for a fixed ARL^0 (Frisé (2003)).

Secondly, an important specification of utility is that of Girshick and Rubin (1952) and Shiryaev (1963). The gain of an alarm is a linear function of the difference $t_A - \tau$ between the time of the change and the time of the alarm. The loss of a false alarm is a function of the same difference. The criterion of maximization of this utility is named the ED criterion, since the expected delay from a change to the detection is minimized. This criterion will be further discussed in Section 5.2, where the exact definition is presented. Their solution to the minimization of the expected utility is identical to the LR method described by Frisé and de Maré (1991). The LR method will be used as a benchmark when comparing the expected delay for EWMA and different modifications in Section 5.2.

The third criterion is the minimax of the expected delay after a change with respect to the time of the change. It will be further discussed in Section 5.3, where the exact definition is presented. It is related to the ED criterion as several possible change times are considered. However, instead of an expected value, which requires a distribution of the time of change, the worst value is used. Thus, minimax solutions, with respect to τ , avoid the requirement of information about the distribution of τ . Important results on minimax are given by e.g. Pollak (1985). The maximal value of the expected delay is for $\tau=1$ for many methods and with a minimax perspective this can be a motivation for the use of ARL^1 . However, this argument is not relevant for all methods. It will be demonstrated that the argument is not useful for the EWMA method. A still more pessimistic criterion, the “worst possible case”, not only uses the worst value of τ but also the worst history of earlier observations. The merits of studies of this criterion have been thoroughly discussed by e.g. Yashchin (1993). Important results based on this minimax criterion were given by e.g. Moustakides (1986), who concludes that the CUSUM method is optimal based on this criterion.

In the studies of the ED criterion and the minimax criterion a fixed value of the total false alarm probability $P(t_A < \tau)$ is used here, whereas a fixed value of the ARL^0 is used in the ARL

criterion. The relation between these two measures of false alarms is illustrated in Section 5.2. For the minimax criterion ARL^0 is often used in literature, as will be discussed in Section 5.3.

5. OPTIMAL EWMA

The properties of the EWMA methods depend both on the value of λ in the alarm statistic and on the alarm limit. The aim is to achieve a combination of limits and values of λ , which results in a total optimality. For the EWMA method there is a restriction of how the given formula for the alarm limit should be combined with weights in the alarm statistic. When optimal weights and optimal limits can be combined, better exponentially weighted methods are achieved as will be seen below.

An important feature of a method is how the alarm limit depends on the decision time s . The unavoidable false alarms can be allowed early or late. This error-spending strategy is important for the properties of the method. It might appear as if the error-spending by EWMAe is independent of λ since $P(Z_s > L\sigma_{Z_s} | \mu=0) = \Phi(-L)$ is independent of λ . However the correlation between successive Z_s , and thus the error-spending, depends on λ .

In order to investigate the properties of methods according to the different optimality criteria we need estimates of the values of different measures. Numerical approximations involve problems, as discussed in the next section. Here a simulation study was performed to investigate the properties of methods. The cases studied were chosen to be representative for situations in practice but only cases which could be studied with enough exactness by simulations were chosen. A special study, reported in the Appendix, demonstrates that the technique used, and the very large number of replicates, allow safe conclusions.

5.1 Minimal ARL^1 for a fixed ARL^0

5.1.1 Choice of value of the parameter λ .

ARL^1 and ARL^0 are expectations under the assumption that there are equal distributions for all observations under each of the two alternatives. Statistical inference with the aim of discriminating between the alternatives that all observations have the expected value 0, or all have the expected value μ , should by the ancillarity principle not be based on the time of the observation. Thus, one should not give unequal weight to the late and old observations. However, the ARL-criterion must not necessarily agree with generally accepted principles of inference.

Frisén (2003) demonstrated that there exist methods with equal weights for all observations which are ARL optimal. This is another reason to choose equal weights for the EWMA method. To get equal weights to all observations by the EWMA method, λ should approach zero.

The variance for the standardized alarm statistic, Z_w , has the smallest variance when λ approaches zero. As the expected value of Z_w is constant (0 or μ for the conditions for ARL^0 or ARL^1 respectively) the coefficient of variation of the statistic is minimized. This is still another argument for using equal weights if the aim is only to discriminate between these two alternatives.

Figure 1 here

Figure 1 supports the conclusion that λ should approach zero to satisfy the ARL criterion. In that figure the observations for $\lambda=0$ and $\lambda=1$ are based on theory while the others are based on simulations. Also, Chan and Zhang (2000) observed by simulations that the ARL^1 decreases when λ decreases. The suggestion was to impose a restriction that the variance of the run length should be small, which it is not for small values of λ . The conclusion that λ

approaching zero is optimal is avoided by this alternative criterion.

Different computational techniques to numerically approximate the ARL have been suggested. Here, we mention some which have been directly devoted to find “optimal” values of the parameter λ in the EWMA method. For one-sided EWMA, Robinson and Ho (1978) used Edgeworth series expansion. According to Lucas and Saccucci (1990) this approximation is inaccurate for small values of λ . The algorithm does not converge for small values of λ . Crowder (1989) used integral calculations suggested by Crowder (1987). Lucas and Saccucci (1990) suggest that the EWMA statistic should be represented as a continuous-state Markov chain, whose properties can be approximated by a finite-state Markov chain following a procedure similar to that of Brook and Evans (1972). Srivastava and Wu (1993) use a continuous time model and an explicit formula for average run lengths. Continuous time approximations are not good for all situations but correction terms are suggested by Srivastava and Wu (1997).

The results concerning the optimal value of λ , by the numerical approximations in the publications mentioned above, are not in agreement with the theoretical arguments that λ should approach zero to give a statistic which is suitable for the ARL criterion. The explanation for this might be the lack of correspondence between the criterion and principles of inference or the numerical approximations. Besides, the present result is for the one-sided case and the results by other authors mentioned above are for the two-sided case.

5.1.2 The alarm limit.

In the preceding section we found that, for ARL optimality, equal weights for all observations should be considered. Still, we have to construct the optimal alarm limit. We will now study alarm limits for EWMA when $\lambda \rightarrow 0$.

With use of the exact variance $\sigma_{Z_S}^2$ we have the limit:

$$L_{EWMAe} = L (\lambda[1 - (1-\lambda)^{2s}]/(2-\lambda))^{1/2}/[1-(1-\lambda)^s] \rightarrow L/s^{0.5} \text{ as } \lambda \rightarrow 0.$$

The method tends to a repeated likelihood ratio test method (see Siegmund (1985) p 86, 98) as λ approaches zero. For EWMAa the situation is more complicated since L_{EWMAa} tends to infinity for all s , for a fixed L , when λ approaches zero. To keep a fixed ARL^0 , L has to be decreased. The ratio between the limit for a certain s and that for $s=1$, will reflect how the alarm-limit depends on s . The ratio

$$L_{EWMAa}(s)/L_{EWMAa}(1) \rightarrow 1/s \text{ as } \lambda \rightarrow 0.$$

Also the SCUSUM method, which gives an alarm at

$$t_A = \min\{s: \sum_{t=1}^s X(t)/s > L/s\},$$

where L is a constant, has an alarm limit which is proportional to $1/s$. There is an alarm when the average of the observations exceeds L/s .

Since the ARL criterion can be criticized, we will also briefly examine how the EWMA satisfies a modification, the criterion of minimum ARL^1 for a fixed false alarm probability (see Section 5.2). The LCUSUM method, demonstrated by Frisén (2003) to be optimal for this criterion is based on an SPRT, and gives an alarm at

$$t_A = \min\{s: \sum_{t=1}^s X(t)/s > \mu/2 + L/s\},$$

where L is a constant. In Figure 3 in the next section the minimal ARL^1 for a fixed false alarm probability is illustrated by the low expected delay for $\tau=1$. However, the low expected delay

is true only for this value of τ . The properties for a later change are very bad. This is so also for the SCUSUM method but less pronounced. Thus, to go beyond the class of EWMA methods you can get still better ARL-properties, but even worse properties for large values of τ .

Even though none of the EWMA variants have alarm limits which depend on s in the same way as the LCUSUM method when λ approaches zero, the EWMAa approaches the LCUSUM method when μ tends to zero.

It follows from the general theory of SPRT that the LCUSUM method does not have a finite ARL^0 . The same follows for SCUSUM from the theory of Brownian motions and the fact that discrete stopping procedures have larger expected stopping times than the continuous version. When λ approaches zero, the ARL^0 of EWMAa behave as for SCUSUM and will not be finite. The alarm limit of EWMAa converges to that of EWMAe when s increases if the same value of L is used. The alarm limit for Z_s is constant for EWMAa and increases faster with s for EWMAe than for EWMAa. Thus, the ARL^0 tends to infinity also for EWMAe. Exact ARL optimality for EWMA is thus hard to discuss. However, as L decreases to $-\infty$ in order to get a fixed value of ARL^0 , $P(t_A=1)$ increases to 1. The results from the simulations, reported in Figure 1 for EWMAa, illustrates that the ARL-properties will approach the optimal one with $ARL^1=1$ for a fixed value of ARL^0 when $\lambda \rightarrow 0$.

Methods which allocate the power to the first time points, like EWMAe, the Fast Initial Response method by Lucas and Saccucci (1990) or a combination of both and a more direct allocation (Steiner (1999)) will have good ARL^1 properties but worse ones if the change happens later. This is also true for EWMAa with small values of λ as illustrated in Figure 3.

5.2 Minimal expected delay

The ED criterion as described in general terms in Section 4, belongs to a wide class of utility functions. The LR method (Frisén and de Maré 1991), which is based on the full likelihood ratio maximizes all these utility functions.

Instead of using the ARL^0 as the false alarm measure as in the ARL criterion we now use a fixed value of the false alarm probability, $P(t_A < \tau)$. The relation between the ARL^0 and the $P(t_A < \tau)$ is illustrated in Figure 2 for some methods. It is clear that you can expect different results in comparisons depending on which of the restrictions you choose. Methods are favored differently by the two restrictions. The EWMAa method with a large value of λ is favored by the ARL restriction, while the LR method (to be described below) optimized for small values of ν and μ is favored by the restriction on $P(t_A < \tau)$.

Figure 2 here

Let the expected delay from the time of change, $\tau=i$, to the time of alarm, t_A , given the time of change, be denoted by

$$ED(t) = E[\max(0, t_A - t) \mid \tau=t]$$

To connect with the preceding section, it can be noted that $ED(1) = ARL^1 - 1$. For most methods the $ED(t)$ will tend to zero as τ increases. The conditional expected delay

$$CED(t) = E[t_A - t \mid \tau=t, t_A \geq t] = ED(t) / P(t_A \geq t)$$

on the other hand, will for most methods converge to a constant value. The $CED(\tau)$ for the EWMA methods, with large values of λ , is fairly independent of τ . Only the last observations will have any great influence. Thus the ability to detect a change, given that no alarm has

been given earlier, is fairly constant. For $\lambda=1$ the EWMA methods equal the Shewhart method, which has a constant $CED= ARL^1-1$. However, for small values of λ the expected delay for changes that occur in the beginning or later on differ much. In Figure 3 the relation between CED and τ is illustrated for some methods described above or below.

Figure 3 here

The summarizing expected delay

$$ED= E_{\tau}[ED(\tau)],$$

where the expectation is with regard to the distribution of the time τ of the change, is in focus of this section. The ED is minimized (for a fixed value of the false alarm probability) by the LR method. For the situation specified in Section 2, the LR method gives an alarm for

$$t_A = \min\{s: \sum_{t=1}^s P(\tau=t)\exp\{t\mu^2/2\} \exp\{\mu \sum_{u=t}^s X(u)\} > \exp\{(s+1)\mu^2/2\} P(\tau>s) \frac{K}{1-K}\}$$

where K is a constant which determines the probability of false alarm. The alarm criterion can equivalently be expressed by the posterior probability

$$P(\tau \leq s | X_s = x_s) > K .$$

The special data-analytic interpretation of the values of the statistics of methods with a simple relation to the posterior probability is pointed out by Kenett and Pollak (1996).

The LR method is optimized for the values of μ and v used in the alarm statistic. When the values of these two parameters are of interest, they are used as arguments in the notation $LR(\mu, v)$.

The expected delays for the LR method and the Shewhart method are presented in Figure 4 for some situations. The expected delay of the LR method is used as the benchmark in the following descriptions, since it is the minimal one. The values for the Shewhart method indicate a practical upper limit, even though EWMA with a very small value of λ has worse values. Thus, the limits between which methods vary for a specific combination of v and $P(t_A < \tau)$ can be seen in the figure. For small values of μ the intervals are wide, while they are tight for large values. The LR and Shewhart methods converge to have identical properties when μ tends to infinity, as proved by Frisén and Wessman (1999).

Figure 4 here

Linear approximations of the LR method were suggested by Frisén (2003). Here we will study three variants and compare them to the EWMA method. The approximations of the LR method are functions of

$$\lambda^* = 1 - \exp(-\mu^2/2)/(1-v),$$

which has a specific value as soon as μ and v are specified. However, here we are specially interested in the methods as modifications of the EWMA method. We will express the methods with the unspecified parameter λ and examine whether the value $\lambda = \lambda^*$ actually is optimal or if it can be improved.

The first approximation, which is denoted LinLR is achieved by a Taylor approximation of the alarm statistic of the LR method. The LinLR method has a linear alarm statistic with the standardized weights

$$w_{\text{LinLR}}(s,t) = \left[\frac{1}{(1-\lambda)^t} - 1 \right] \frac{\lambda(1-\lambda)^s}{1 - (1+s\lambda)(1-\lambda)^s}.$$

The weights are well approximated by exponential weights proportional to $1/(1-\lambda)$ except possibly for small values of t . The alarm limit is

$$L_{\text{LinLR}}(s) = \frac{\lambda [v(1-L)(1-\lambda)^{s+1} - 1 + \lambda L]}{v(1-L)(1-\lambda)\{\mu - \mu(1+s\lambda)(1-\lambda)^s\}},$$

where the constant L is determined by false alarm properties. This alarm limit, as well as that of the normalized version of the EWMAa, tends to a constant when s increases. The LinLR method will give an alarm for

$$t_A = \min\left\{s: \sum_{t=1}^s w_{\text{LinLR}}(s,t)X(t) > L_{\text{LinLR}}(s)\right\}.$$

The exponential weights of the EWMA are appealing. When the weights of the LinLR method are approximated by exponential weights, and we keep the limit of the LinLR method, we have the EWLRL method, which gives an alarm at

$$t_A = \min\{s: ZW_s > L_{\text{LinLR}}(s)\}.$$

A third approximation, EWInLR, is achieved by a Taylor approximation of the logarithm of the alarm statistic of the LR method and, as a further approximation, exponential weights. An alarm is given at

$$t_A = \min\{s: ZW_s > L_{\text{EWInLR}}(s)\},$$

where

$$L_{\text{EWInLR}}(s) = \frac{\lambda [1 - (1-\lambda)^s] [L - \ln(1 - (1-\lambda)^s)]}{\mu [1 - (1+s\lambda)(1-\lambda)^s]}.$$

The three approximations of the LR method can be seen as modifications of the EWMA method. The EWLR and EWInLR methods keep the alarm statistic of the EWMA but modify the alarm limits. The modifications are negligible for large values of the decision time, s but have an influence by the effect for small values of s . The LinLR method modifies also the alarm statistic for small values of s . The modifications of the EWMA have the purpose to give smaller expected delays.

The CED is illustrated in Figure 3. For the EWInLR method it can be seen that the dependency of CED on τ is very similar to that for LR. The other LR-approximations are not illustrated in the figure but are also very similar to the LR method.

The methods are compared in Figure 5 by the ED relative to that of the LR method for different values of λ for some situations. In this figure we can find the details, discussed in Section 5.2.1, about how the value of λ influences the ED for different methods. The diagrams are also used for derivation of the summarizing measures, used to compare the methods in Section 5.2.2.

Figure 5 here

5.2.1 Optimal λ .

The choice of λ is important and the search for the optimal value of λ has been of great interest in literature, as discussed in Section 5.1.1 concerning the ARL criterion. In Figure 3 it is demonstrated how different values of λ influence the ability of detection at different time points. Small values of λ result in good ability to detect early changes while larger values are necessary for changes that occur later. This illustrates the importance of the choice of λ also for the ED criterion.

The minima of ED with respect to λ are rather flat as demonstrated in Figure 5. The minimal value of ED can thus be determined accurately, but there is an uncertainty on the

corresponding λ . This also means that an exact value of the optimal value of λ is not very important. However, an approximate value is important since the optimal value differs much between situations and the ED can be very large far away from the optimal value. The values of λ which give the minima of ED with respect to λ are determined by quadratic interpolation for the neighboring observed values of λ . The optimal value of λ increases with μ and decreases with v .

The formula for λ^* is suggested to achieve approximations of LR. Since this formula is a result of several approximations, it can only be expected to give rough approximations of the optimal values of λ for different situations. However, as can be seen in the upper row of Figure 6, the formula results in very good properties for the modifications (within 5% of the value of the LR-method) for most cases studied. Comparison of the upper and lower row of Figure 6 reveals that the loss by use of λ^* instead of the exactly optimal value of λ is small and less than 5% for the modifications, for all cases studied. This is also true for the EWMAa for large changes, which are the only cases where EWMAa should be used, as will be demonstrated below. Thus, the formula is useful for the choice of value of λ .

Figure 6 here

5.2.2 Comparison between methods.

In this section we compare how well the methods perform when the optimal value of λ for each method and situation is used. These minimal ED values are presented in the bottom row of Figure 6. They are presented in relation to the minimal possible values which are achieved by the LR method.

All the methods studied perform considerably better than the Shewhart method, except for large changes, where the differences are less. The Shewhart method is not included in Figure 5 and 6 since the values are too high to be included in the scale suitable for the other methods.

The EWLR method is very good for small values of μ . The minimal value of ED is within 8% of that for LR for all cases studied, and for most cases much smaller. The LinLR method could be expected to be better than the EWLR method since the approximation of the weights is avoided. This approximation could be expected to have most effect for small values of μ where small values of λ are used. However, the improvement over EWLR is slight. Thus, there is no indication of a need to modify the EWMA alarm statistic.

EWlnLR has the smallest values of ED, except for very small values of μ , where all the modifications are very close to the LR method. The minimal value of ED is within 5% of LR for all cases studied.

The EWMAa method is very bad for small changes. However, for large values of μ EWMAa is slightly better than the EWLR method and approximately equal to the EWlnLR method.

The three different cases studied have different false alarm properties. The largest ARL^0 is for $v = 0,1$ and $P(t_A < \tau) = 0,1$. The smallest is for $v = 0,01$ and $P(t_A < \tau) = 0,75$.

For the cases studied, all methods (with the optimal value of λ) have expected delays between the LR and Shewhart methods. The intervals between these two are illustrated in Figure 4.

5.3 Minimax

Examples of $CED(\tau)$ curves are given in Figure 3 where it is obvious that for EWMA the maximal value is not always for $\tau=1$. Thus, ARL^1 does not reflect the maximal value. The maximal $CED(\tau)$ with respect to τ was first noted and then the minimal value with respect to λ was determined. These minimax values of λ and also the optimal values with respect to the ED criterion are given for some cases in Figure 7. It is clearly seen that the values of λ by the two criteria are closely related. This is not at all the case for the ARL optimality where $\lambda=0$ is

the solution. Thus, the optimal value of λ for the minimax and the ED optimality are related but the minimax and ARL optimality differ much.

The Shewhart method, which is very bad by the other criteria, is nearly as good as the LR method by the minimax criterion. The EWMA method with a value of λ near λ^* is the best of the examined methods by this criterion.

Most minimax studies in the literature are for a fixed ARL^0 and not for fixed $P(t_A < \tau)$ as here. This makes a difference as is seen in Figure 2. However, the slight difference in the present situation is in a direction to strengthen the above conclusions.

Figure 7 here

6. DISCUSSION

There are two necessary conditions for a method to satisfy an optimality criterion. One is that the different observations at every fixed decision time, s , have optimal relative weights. The other one is that the alarm limit, as a function of s , is optimal. The alarm limit determines if the unavoidable false alarms will be spent early or late. For the conventional EWMA, the fact that both the weights and the alarm limit are completely determined by the parameter λ hampers the possibility to achieve both at the same time. This is true, both for the criterion of minimal ARL^1 and the criterion of minimal expected delay. In earlier attempts to achieve an optimal EWMA the alarm limit has not been questioned and the optimal weighting has been determined given this function of s . The idea of a constant probability of the EWMAe statistic to exceed the alarm limit might seem appealing. However, for surveillance, there is no special merit with this. Besides, the asymptotic version completely destroys this property for small values of λ . Here we examine both the weights and the limits.

It is demonstrated that the ARL criterion can be questioned as a formal criterion. Sometimes it is claimed that the ARL criterion is enough since no great differences between

the properties for different values of τ exist. This is true for very large values of λ , but certainly not for all values of λ . For the detection of small changes (e.g., $\mu=0.5$), and the small values of λ of EWMA suitable for that case, the conditional expected delay is strongly dependent on the time of the change. This is illustrated in Figure 3. In this figure it is also demonstrated that the ARL criterion differs much from the minimax criterion. We suggest that when immediate changes are the main interest, methods for sequential testing of hypotheses such as the SPRT-based LCUSUM method, and not EWMA, should be used.

For the ED criterion it is necessary with some knowledge of the parameter v in the distribution of the time of the change. In practice some knowledge should be available and should influence the choice of method. However, exact knowledge is rarely available. Frisé and Wessman (1999) demonstrated that the LR method is very robust with regard to v , (no major effects were demonstrated for v less than 0.1). The lack of need of a distribution for the change point is an advantage for the ARL- and minimax optimality. However, the robustness of the LR method makes this disadvantage of the ED criterion less important. Besides, it is not self evident that the possibility to optimize a method for a parameter should be seen as a disadvantage even though it is hard to specify which value is of most interest.

The necessity to find the (approximately) optimal value of the parameter λ in order to get good properties is a problem. For the ED criterion the suggested formula for λ^* gives good values of the expected delay for all methods and cases studied here. The only exception is for the EWMAa for very small changes, where the EWMAa should not be used anyway. When conclusions are based on simulations, it is always a question about the generality of the results. As indicated above the results on the usefulness of the approximations of the LR method could not be guaranteed for extreme situations but the cases studied represent reasonable variation in values of μ and v of interest for practical applications.

The EWMAa can be improved, with respect to the expected delay, by modifying the alarm

limit to that of EWlnLR. The improvement is substantial for small changes, where it can be even slightly more improved by EWLR or LinLR. The LR method is of course best with respect to expected delay, but the approximations are nearly as good.

For large values of μ there is no great difference between the EWMAa and the other methods. Thus, for large values of μ there is no need for the modifications. However, the modifications have clearly less expected delay than the EWMAa for small values of μ .

EWMAa with $\lambda = \lambda^*$ approximates the LR-method. Many of the CED-curves for this choice of λ are rather flat and this adds to the explanation of the similarity between the minimax solution and the ED solution for EWMA. ARL optimality requires a small value of λ and then the steep slope of the CED-curves for these small values explains the lack of correspondence between the minimax- and ARL optimality for the EWMA method.

APPENDIX. Accuracy of the results from the simulation study.

When conclusions are based on simulations it is necessary to ensure that the number of replicates is enough. In many studies, including this one, different methods are evaluated with respect to some measure, θ under the restriction that another measure, P , has a specified value. In this appendix we describe the statistical technique used for calculation of confidence intervals of the results from the simulations and also exemplify the accuracy of the results displayed in the figures. The measure θ depends on the out of control run length distribution, which in turn makes θ a function, $\theta(L)$, of the constant L used in the alarm limit. The measure θ could for example be the ARL^1 , CED or ED for different methods. The measure P , which is specified to P^* for comparability is here ARL^0 or $P(t_A < \tau)$. Also this measure depends on the value of the alarm limit. The aim is to study the accuracy of the estimates of θ for $P = P^*$. By analyzing the accuracy in each step used to produce the results, and combining these, we get conservative confidence intervals for the points in the figures.

We start by studying the accuracy of the value of L . Denote by L^* the value of L such that $P(L^*) = P^*$ and by L' the value of L estimated by our procedure and for which θ is estimated. The procedure used here to estimate L^* is to choose L_1, L_2, \dots, L_n and use simulations for each value of L to estimate the values of $P(L_1), P(L_2), \dots, P(L_n)$. We approximate $P(L)$ with a linear function locally, $\hat{P}(L) = \hat{a}_1 + \hat{b}_1 L$, and choose L' accordingly to give $\hat{P}(L') = P^*$. A confidence interval CI^1 for L^* (and thus also for $L^* - L'$) can be constructed as consisting of those values of L which would not be rejected by the test of $H_0 : P(L) = P^*$. The very large number of replicates ensures that the normal distribution can be used.

A confidence interval CI^2 for $\theta(L')$ can be constructed using simulations for the chosen value L' . Also in this case we can use a normal approximation due to the large number of

replicates.

The last link is to determine how influential the error in L is with respect to the value of θ . Estimates of $\theta(L)$ for some values L_1, L_2, \dots, L_n of L around L' are achieved by special simulations for the purpose of determining the accuracy. We approximate $\theta(L)$ locally by a linear function $\theta(L^*) = \theta(L') + b_2(L^* - L')$. A confidence interval CI^3 for b_2 in the regression can be constructed.

The confidence intervals from the three steps described above can then be combined to form a confidence interval for $\theta(L^*)$. Let the confidence interval CI^1 for $L' - L^*$ be of confidence $(1 - \alpha_1)$, the confidence interval CI^2 for $\theta(L')$ be of confidence $(1 - \alpha_2)$ and the confidence interval CI^3 , for b_2 be of confidence $(1 - \alpha_3)$. Then, we can combine these intervals to construct a confidence interval of confidence at least $(1 - \alpha_1)(1 - \alpha_2)(1 - \alpha_3)$ for $\theta(L^*)$ by taking $\min\{a + b \cdot c; a \in CI^2, b \in CI^3, c \in CI^1\}$ to be the lower limit and $\max\{a + b \cdot c; a \in CI^2, b \in CI^3, c \in CI^1\}$ to be the upper. We choose $1 - \alpha_1 = 1 - \alpha_2 = 1 - \alpha_3 = 0.983$ which results in a confidence interval for $\theta(L^*)$ of confidence at least 95%. Confidence intervals for $\theta(L^*)$ constructed in this way for some cases studied in the paper will now be exemplified.

Accuracy of the results in Figure 1: Figure 1 shows ARL^1 as a function of λ for the EWMAa method when $ARL^0 = 100$. We exemplify the accuracy in Table A.1. For the determination of the alarm limit L' we have used 500.000 replicates for 25 neighbouring values of L . The confidence intervals for b_2 are based on 10 values of L with 10.000.000 replicates each. The estimates of $\theta(L')$ are based on 1.000.000 replicates. The conclusion that λ should approach zero in order to minimize ARL^1 when $ARL^0 = 100$ is thus well supported. The width of the confidence interval is short compared with the thickness of the lines in

Figure 1.

Table A.1. Examples of the components for the confidence intervals for different values of Figure 1.

λ	Confidence interval for $L^* - L'$		Confidence interval for b_2		Confidence interval for $ARL^1(L')$		Confidence interval for $ARL^1(L^*)$	
	lower	upper	lower	upper	lower	upper	lower	upper
0.001	-0.00022	0.00029	18.75	21.73	2.00	2.02	2.00	2.02
0.010	-0.00044	0.00045	7.07	8.08	4.62	4.64	4.61	4.64
0.100	-0.00055	0.00058	2.42	4.49	5.64	5.66	5.64	5.67

Accuracy of the results in Figure 3: Figure 3 shows CED as a function of τ for the case when $\mu = 0.25$, $\nu = 0.01$ and $P(t_A < \tau) = 0.75$. As an example we calculate a 95% confidence interval for CED(1) for the LR method. For this case the determination of the alarm limit L' was based on 500.000 replicates for 24 neighbouring values of L . This resulted in a confidence interval for $L^* - L'$ of (-0.06, 0.04). Again, the confidence interval for b_2 was based on 10 values of L with 10.000.000 replicates each. This results in a confidence interval of (0.39, 0.44) for b_2 . A confidence interval of (17.23, 17.28) was constructed using 1.000.000 replicates for the estimate of CED(1) using the value L' in the alarm limit. The width of the resulting 95% confidence interval for CED(L^*), (17.2067, 17.2887), is small compared with the line width in Figure 3.

Accuracy of the results in Figure 4: For the same case as exemplified for Figure 3 we can construct a conservative confidence interval for

$$ED(L^*) = \sum_{t=1}^{\infty} P(\tau = t) ED(t, L^*)$$

by using the confidence intervals calculated for ED(t) by the technique given for CED(t) above. Under the assumption that all the steps given above result in an approximately normally distributed estimator we can estimate the variance for the estimator of ED(t). The

variance of the weighted sum $ED(L^*)$ has less variance than the component with the largest variance. The simulations indicate that this is the case for $\tau=1$. Since $ED(1) = CED(1)$, we can use the results from the example above and we have a 95% confidence interval for the ED for the LR method as (2.4976, 2.5796). The width of this confidence interval is small compared with the line width in Figure 4.

The confidence intervals for the results in this paper are small enough using the chosen number of replicates to ensure that the results are reliable. In most cases the confidence intervals are shorter than the thickness of the lines in the figures.

REFERENCES

- Andersson, E. (2002), "Monitoring Cyclical Processes - a Nonparametric Approach," *Journal of Applied Statistics*, 29, 973-990.
- Bodden, K. M., and Rigdon, S. E. (1999), "A Program for Approximating the In-Control ARL for the MEWMA Chart," *Journal of Quality Technology*, 31, 120-123.
- Brook, D., and Evans, D. A. (1972), "An Approach to the Probability Distribution of Cusum Run Length," *Biometrika*, 59, 539-549.
- Carlyle, W. M., Montgomery, D. C., and Runger, G. C. (2000), "Optimization Problems and Methods in Quality Control and Improvement," *Journal of Quality Technology*, 32, 1-19.
- Champ, C. W., Woodall, W. H., and Mohsen, H. (1991), "A Generalized Quality Control Procedure," *Statistics & Probability Letters*, 11, 211-218.
- Chan, L. K., and Zhang, J. (2000), "Some Issues in the Design of Ewma Charts," *Communications in Statistics. Simulations and Computations*, 29, 207-217.
- Crowder, S. (1987), "A Simple Method for Studying Run-Length Distributions of Exponentially Weighted Moving Average Charts," *Technometrics*, 29, 401-407.
- Crowder, S. V. (1989), "Design of Exponentially Weighted Moving Average Schemes," *Journal of Quality Technology*, 21, 155-162.
- Domangue, R., and Patch, S. C. (1991), "Some Omnibus Exponentially Weighted Moving Average Statistical Process Monitoring Schemes," *Technometrics*, 33, 299-313.
- Frisén, M. (1992), "Evaluations of Methods for Statistical Surveillance," *Statistics in Medicine*, 11, 1489-1502.
- Frisén, M. (in press), "Statistical Surveillance. Optimality and Methods.," *International Statistical Review*.
- Frisén, M., and de Maré, J. (1991), "Optimal Surveillance," *Biometrika*, 78, 271-280.

- Frisén, M., and Wessman, P. (1999), "Evaluations of Likelihood Ratio Methods for Surveillance. Differences and Robustness.," *Communications in Statistics. Simulations and Computations*, 28, 597-622.
- Gan, F. F. (1995), "Joint Monitoring of Process Mean and Variance Using Exponentially Weighted Moving Average Control Charts," *Technometrics*, 37, 446-453.
- Gan, F. F. (1998), "Designs of One- and Two-Sided Exponential Ewma Charts," *Journal of Quality Technology*, 55-69.
- Girshick, M. A., and Rubin, H. (1952), "A Bayes Approach to a Quality Control Model.," *The Annals of Mathematical Statistics*, 23, 114-125.
- Kemp, K. W. (1961), "The Average Run Length of the Cumulative Sum Chart When a V-Mask Is Used," *Journal of the Royal Statistical Society Ser. B*, 23, 149-153.
- Kenett, R., and Pollak, M. (1996), "Data-Analytic Aspects of the Shiryaev-Roberts Control Chart: Surveillance of a Non-Homogeneous Poisson Process," *Journal of Applied Statistics*, 23, 125-137.
- Lai, T. L. (1995), "Sequential Change-point Detection in Quality-Control and Dynamical-Systems," *Journal of the Royal Statistical Society Ser. B*, 57, 613-658.
- Lowry, C. A., and Montgomery, D. C. (1995), "A Review of Multivariate Control Charts," *IIE Transactions*, 27, 800-810.
- Lu, C. W., and Reynolds Jr, M. R. (1999), "Ewma Control Charts for Monitoring the Mean of Autocorrelated Processes," *Journal of Quality Technology*, 31, 166-188.
- Lucas, J. M., and Saccucci, M. S. (1990), "Exponentially Weighted Moving Average Control Schemes: Properties and Enhancements," *Technometrics*, 32, 1-12.
- Margavio, T. M., Conerly, M. D., Woodall, W. H., and Drake, L. G. (1995), "Alarm Rates for Quality-Control Charts," *Statistics & Probability Letters*, 24, 219-224.

- Moustakides, G. V. (1986), "Optimal Stopping Times for Detecting Changes in Distributions," *The Annals of Statistics*, 14, 1379-1387.
- Ng, C. H., and Case, K. E. (1989), "Development and Evaluation of Control Charts Using Exponentially Weighted Moving Averages," *Journal of Quality Technology*, 21, 242-250.
- Pettersson, M. (1998), "Monitoring a Freshwater Fish Population: Statistical Surveillance of Biodiversity.," *Environmetrics*, 9, 139-150.
- Pollak, M. (1985), "Optimal Detection of a Change in Distribution," *The Annals of Statistics*, 13, 206-227.
- Prabhu, S. S., and Runger, G. C. (1997), "Designing a Multivariate Ewma Control Chart," *Journal of Quality Technology*, 29, 8-15.
- Roberts, S. W. (1959), "Control Chart Tests Based on Geometric Moving Averages," *Technometrics*, 1, 239-250.
- Robinson, P. B., and Ho, T. Y. (1978), "Average Run Lengths of Geometric Moving Average Charts by Numerical Methods," *Technometrics*, 20, 85-93.
- Shewhart, W. A. (1931), *Economic Control of Quality of Manufactured Product*, London: MacMillan and Co.
- Shiryayev, A. N. (1963), "On Optimum Methods in Quickest Detection Problems," *Theory of Probability and its Applications*, 8, 22-46.
- Siegmund, D. (1985), *Sequential Analysis. Tests and Confidence Intervals.*, New York: Springer.
- Sonesson, C., and Bock, D. (2003), "A Review and Discussion of Prospective Statistical Surveillance in Public Health," *Journal of the Royal Statistical Society Ser.A*, 166, 5-21.
- Srivastava, M. S., and Wu, Y. (1993), "Comparison of Ewma, Cusum and Shiryayev-Roberts Procedures for Detecting a Shift in the Mean.," *Annals of Statistics*, 21, 645-670.

- Srivastava, M. S., and Wu, Y. (1997), "Evaluation of Optimum Weights and Average Run Lengths in Ewma Control Schemes," *Communications in Statistics. Theory and Methods*, 26,1253-1267.
- Steiner, S. H. (1999), "Ewma Control Charts with Time-Varying Control Limits and Fast Initial Response," *Journal of Quality Technology*, 31, 75-86.
- Tsui, K. L., and Woodall, W. H. (1993), "Multivariate Control Charts Based on Loss Functions," *Sequential Analysis*, 12, 79-92.
- Woodall, W. H., and Montgomery, D. C. (1999), "Research Issues and Ideas in Statistical Process Control," *Journal of Quality Technology*, 31, 376-386.
- Yashchin, E. (1993), "Statistical Control Schemes - Methods, Applications and Generalizations," *International Statistical Review*, 61, 41-66.

LEGENDS TO FIGURES

Figure 1. ARL^1 as a function of λ for EWMA when $ARL^0 = 100$ and $\mu=1$.

Figure 2. The false alarm probability $P(t_A < \tau)$ when $ARL^0 = 100$.

Figure 3. The conditional expected delay CED as a function of τ for $\nu=0.01$ and $\mu=0.25$. The false alarm probability is fixed to 0.75.

Figure 4. The expected delay as a function of μ for the Shewhart and LR methods for different values of intensities and fixed false alarm probabilities. The format “Method($P(t_A < \tau)$, ν)” is used for the legend to the curves.

Figure 5. The expected delay relative to the LR method for fixed false alarm probabilities as functions of λ for different variants of EWMA. The values of ν and $P(t_A < \tau)$ are fixed for each column while the values of μ are indicated in each diagram.

Figure 6. The expected delay, relative to LR, for fixed false alarm probabilities as functions of μ for different variants of EWMA. The upper row is for $\lambda = \lambda^*$ while the bottom row is for the optimal value of λ . The values of ν and $P(t_A < \tau)$ are fixed for each column.

Figure 7. The optimal value of the parameter λ of the EWMA method for the ED criterion and for the minimax criterion. The symbol “x” is used for the ED criterion and the symbol “□” is used for the minimax criterion. The dotted curves are for the situation with $P(t_A < \tau) = 0.25$ and $\nu = 0.1$. The solid curves are for the situation with $P(t_A < \tau) = 0.75$ and $\nu = 0.01$.

FIGURE 1

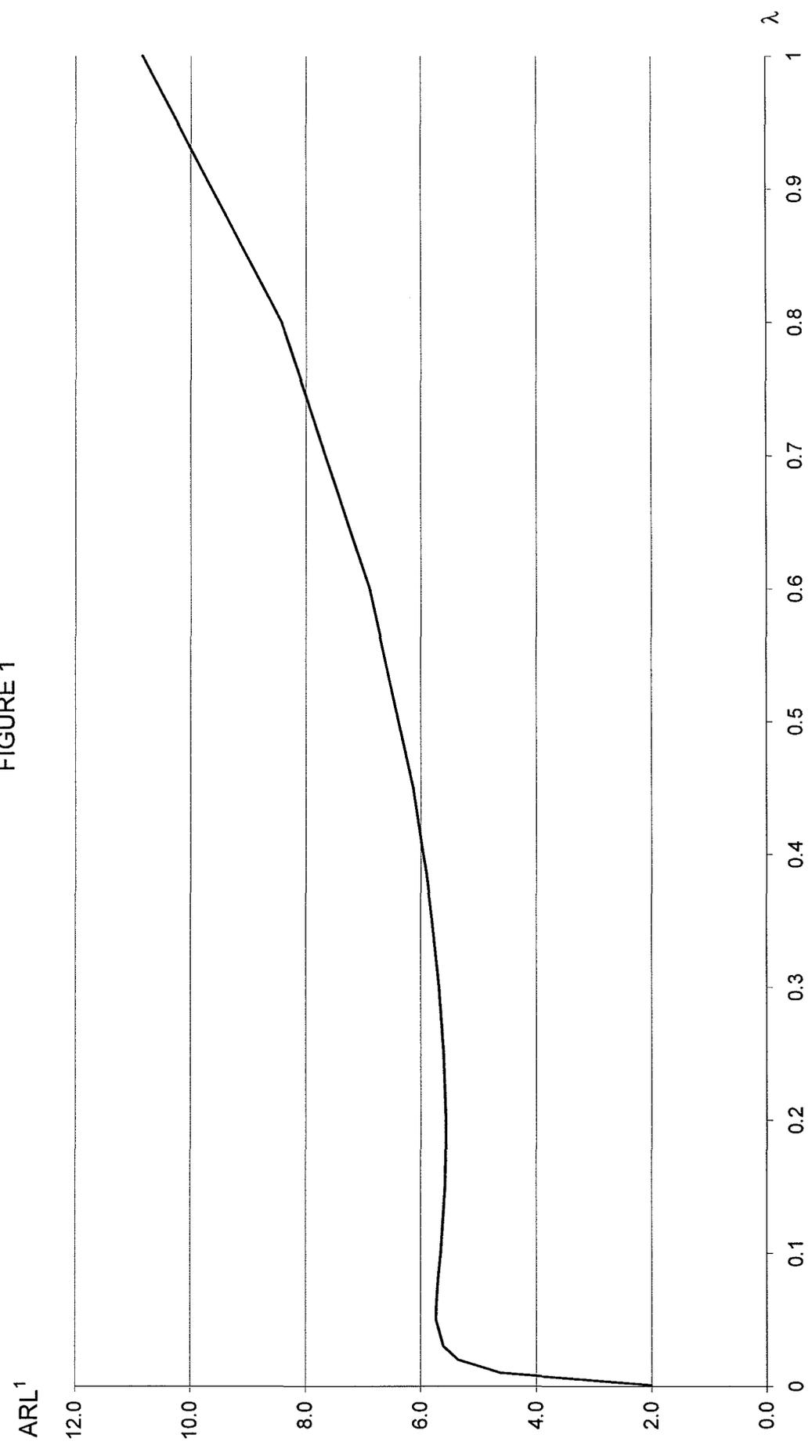


FIGURE 2

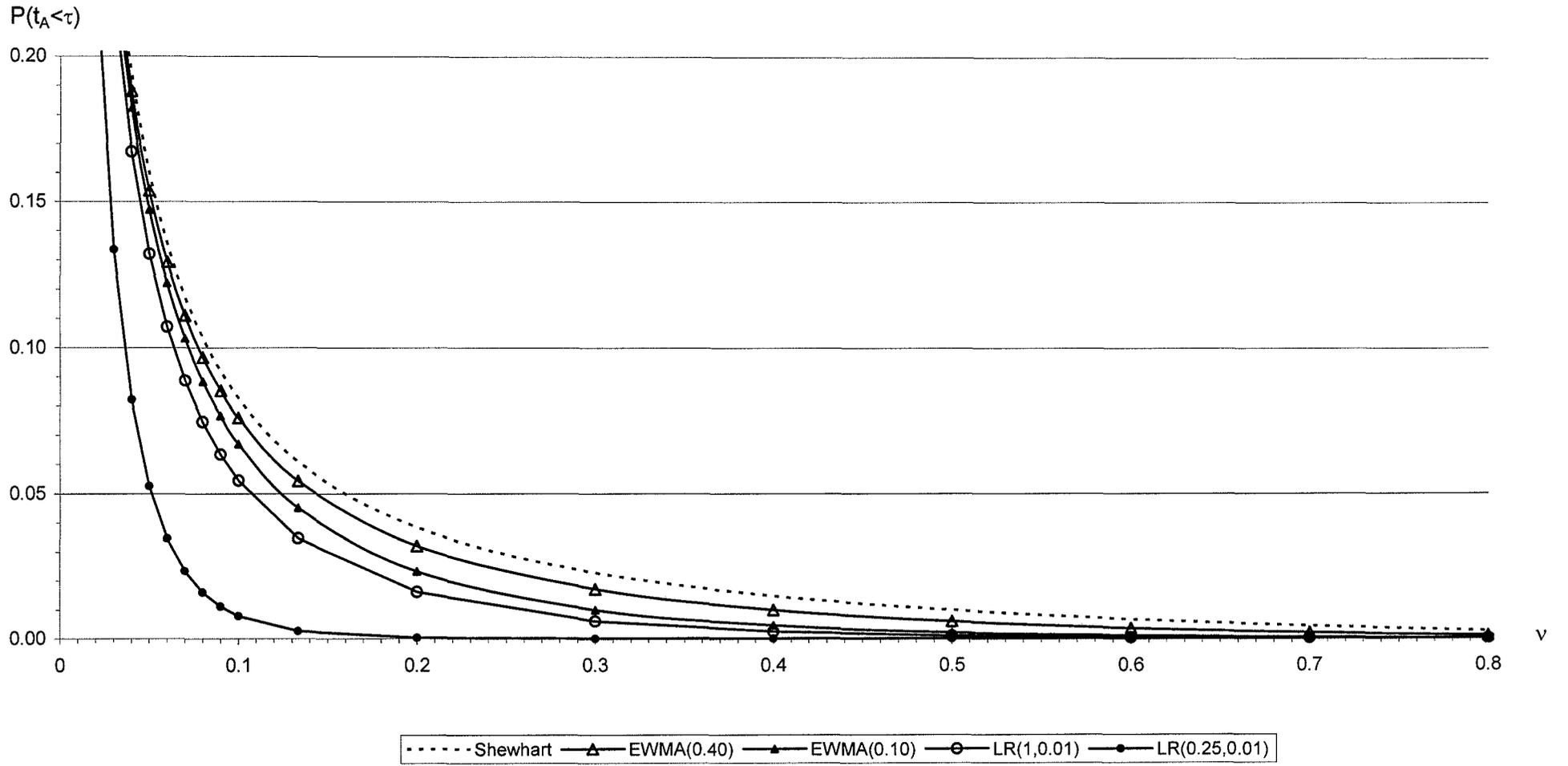
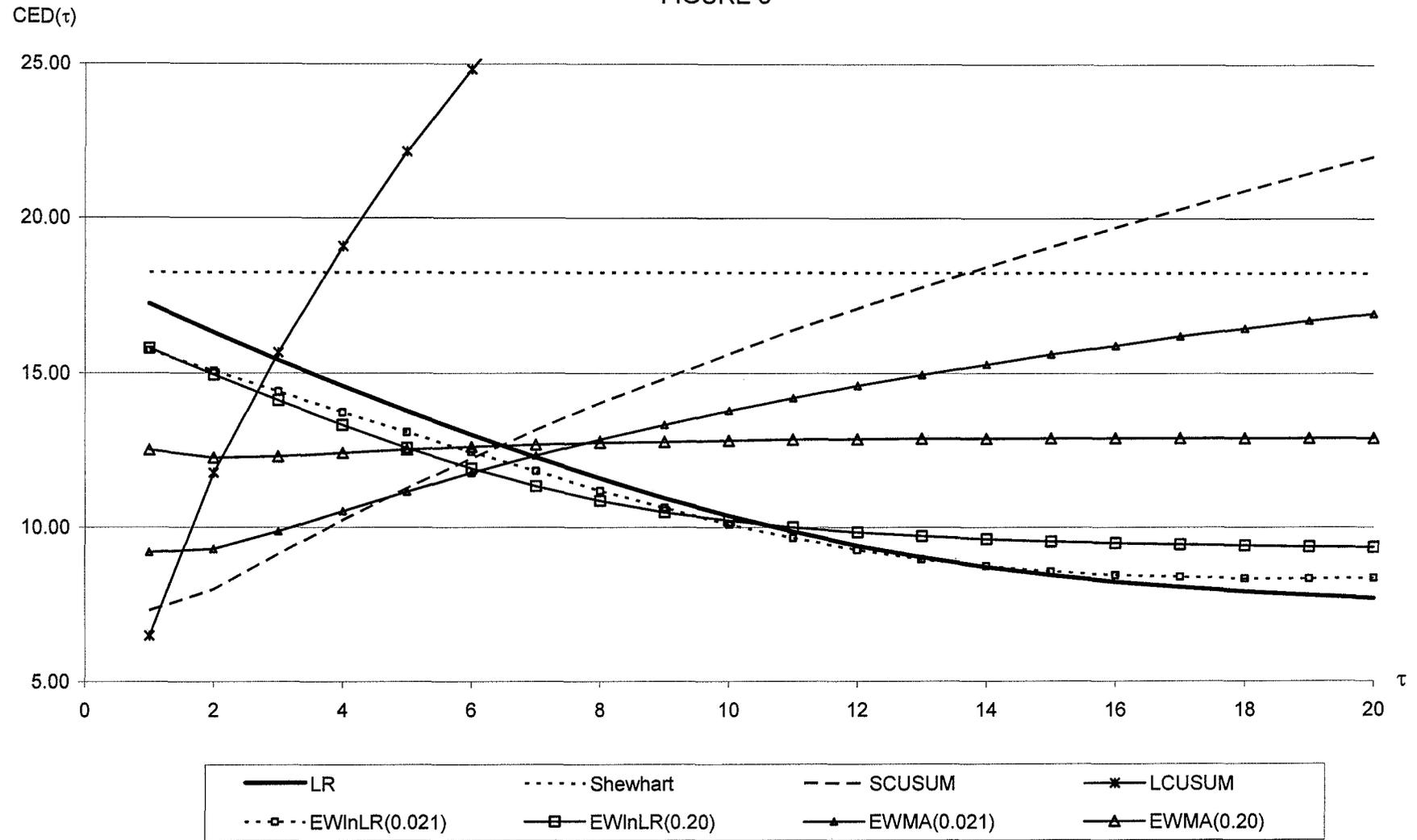


FIGURE 3



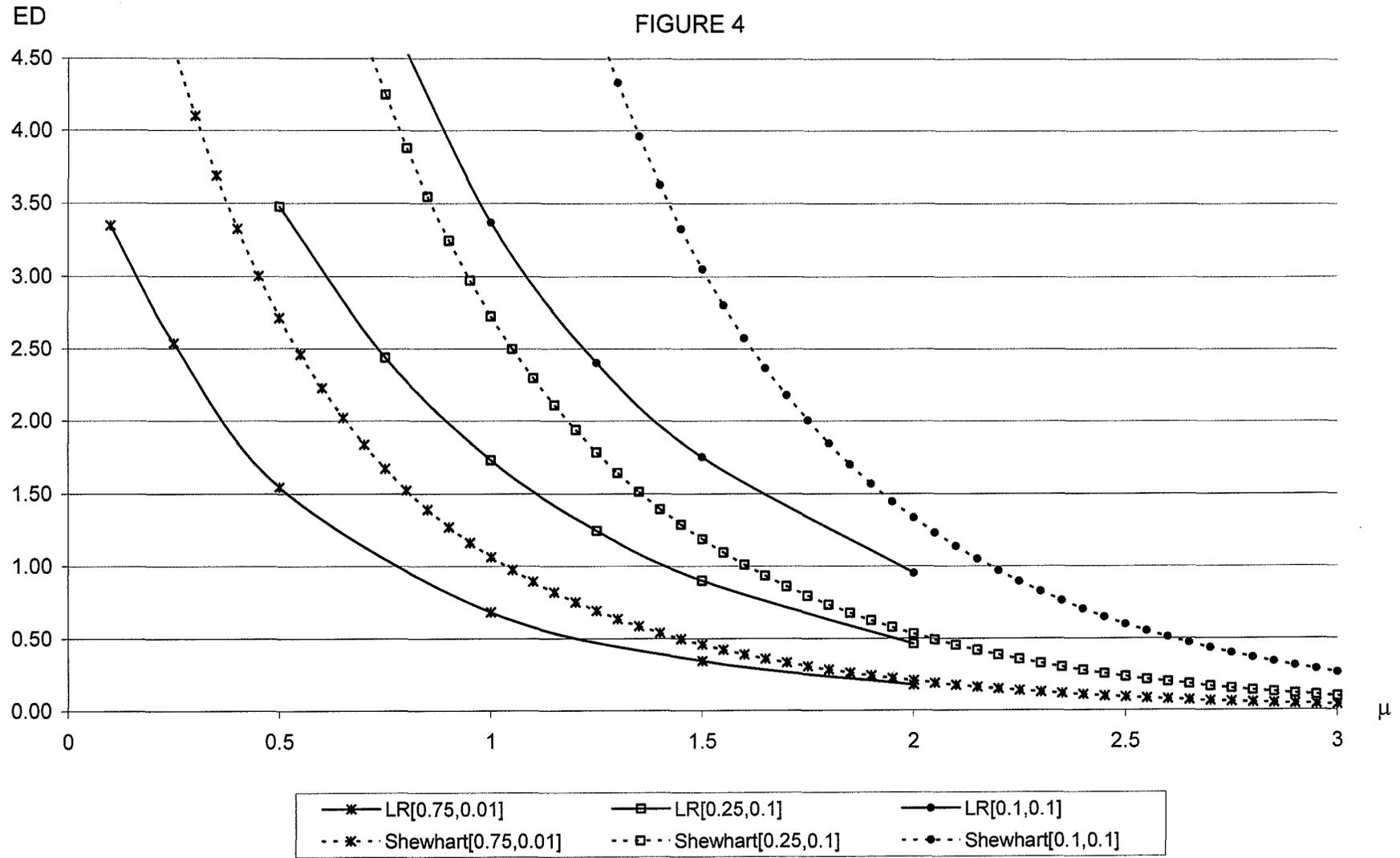


FIGURE 5:

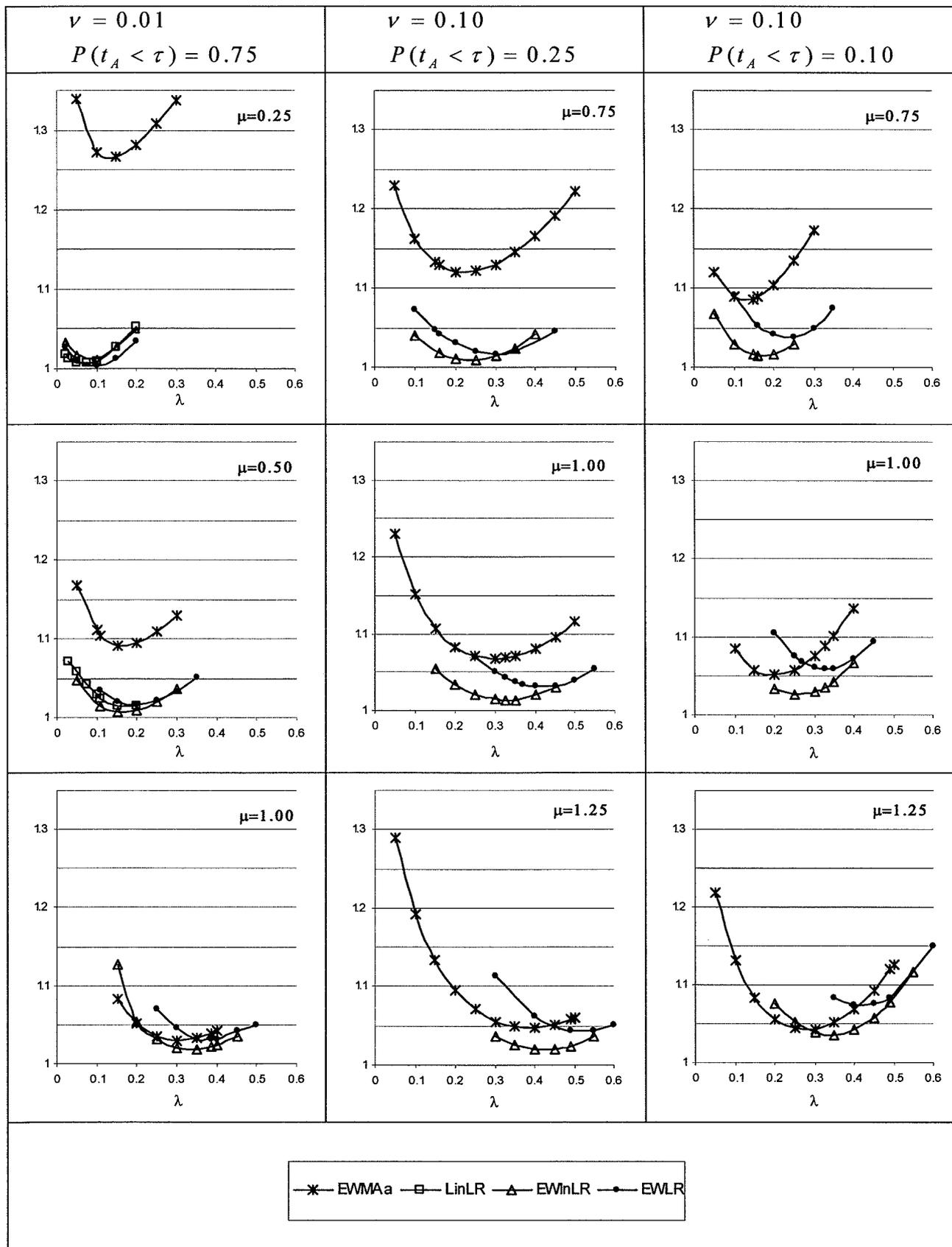


FIGURE 6:

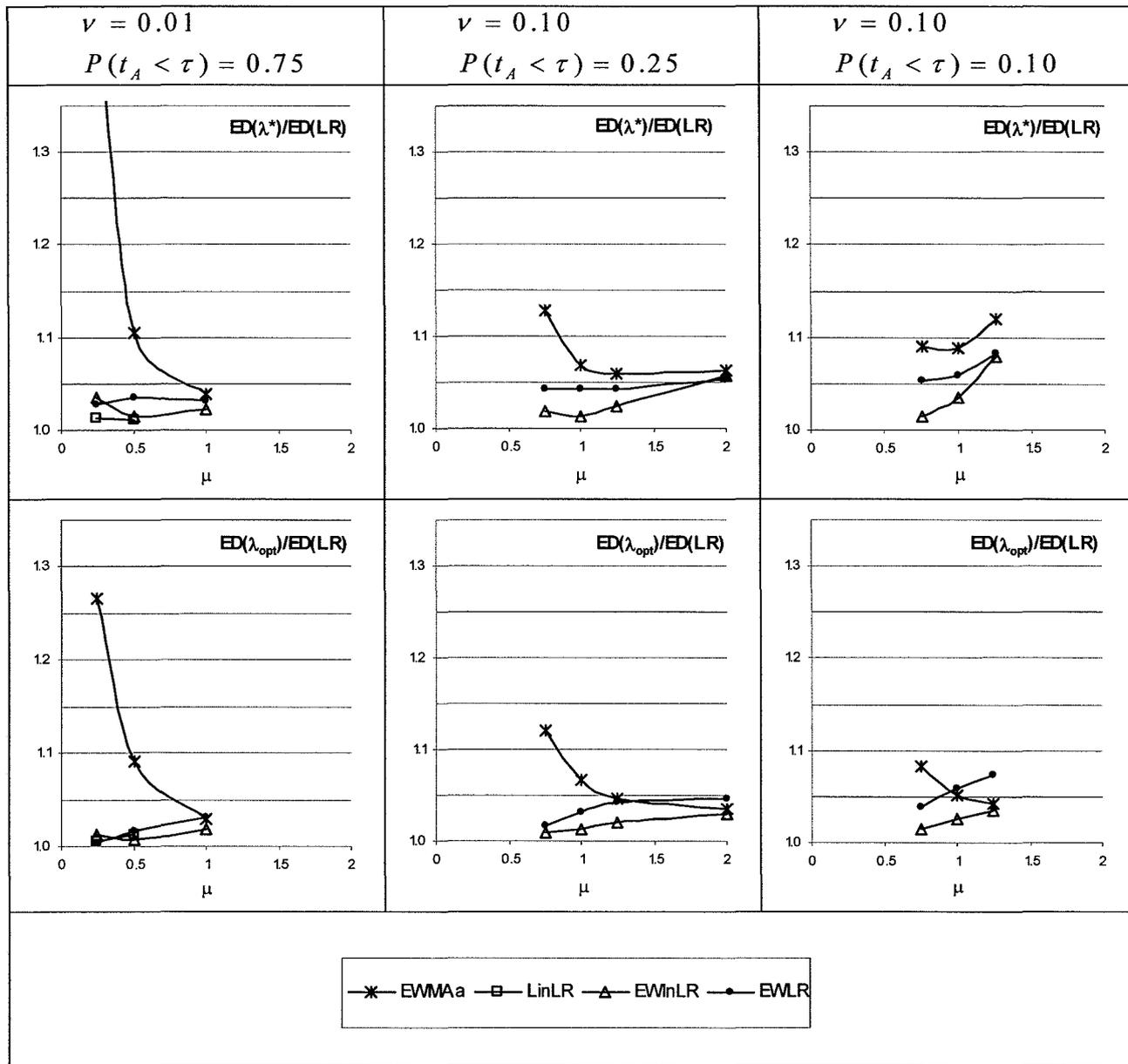
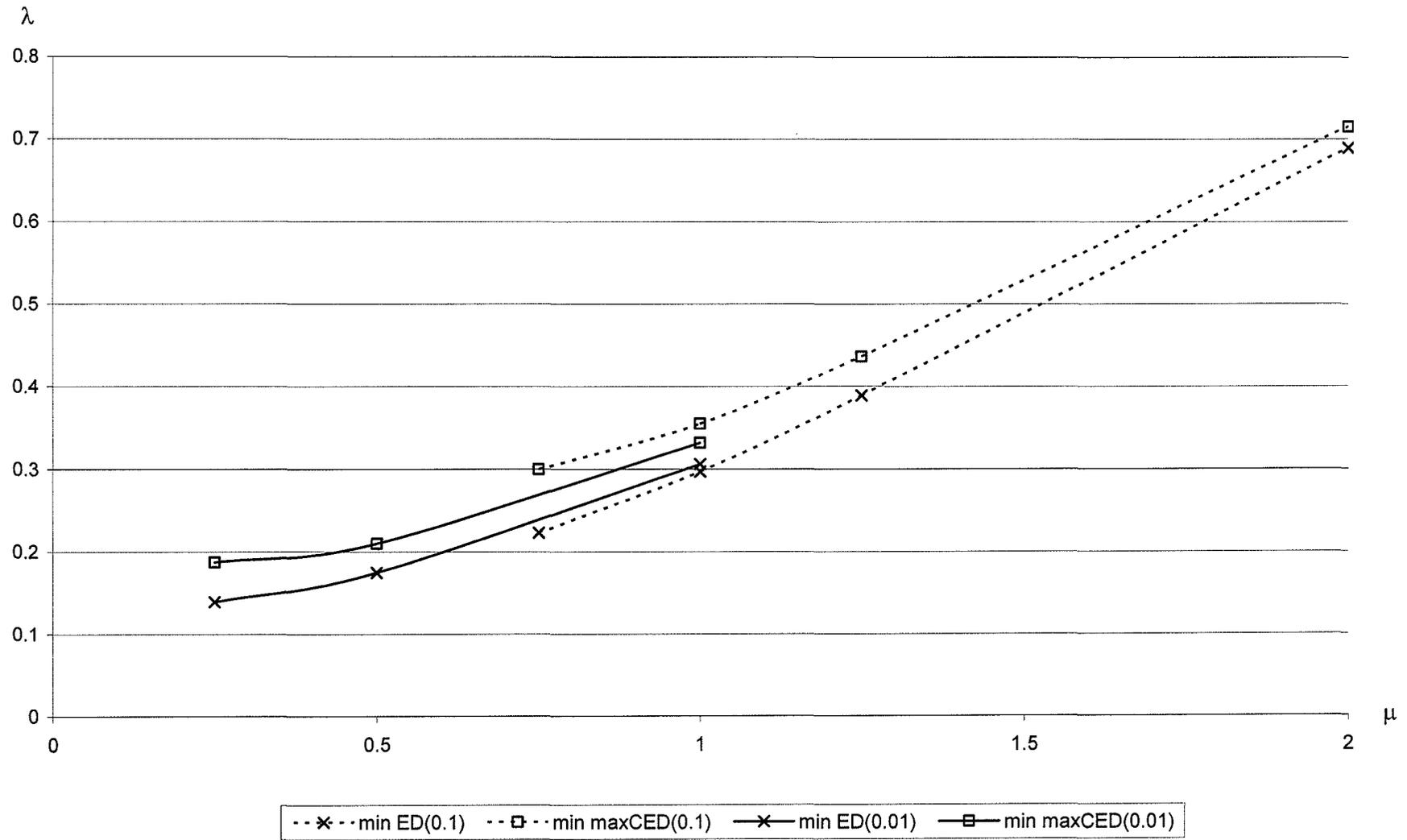


FIGURE 7



Research Report

- | | | |
|--------|---|--|
| 2001:1 | Holgersson, H.E.T.: | On assessing multivariate normality. |
| 2001:2 | Sonesson, C. &
Bock, D.: | Statistical issues in public health monitoring –
A review and discussion. |
| 2001:3 | Andersson, E.: | Turning point detection using non-parametric
statistical surveillance. Evaluation of some
influential factors. |
| 2001:4 | Andersson, E. &
Bock, D.: | On seasonal filters and monotonicity. |
| 2001:5 | Andersson, E.,
Bock, D. &
Frisén, M.: | Likelihood based methods for detection of
turning points in business cycles.
A comparative study. |
| 2001:6 | Sonesson, C.: | Evaluations of some exponentially weighted
moving average methods. |
| 2001:7 | Sonesson, C.: | Statistical surveillance.
Exponentially weighted moving average
methods and public health monitoring. |