Research Report
Department of Statistics
Göteborg University
Sweden

Evaluations of likelihood ratio methods for surveillance.

Differences and robustness.

Marianne Frisén
Peter Wessman

Research Report 1998:2
ISSN 0349-8034

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

INTRODUCTION

In many areas there is a need for continual observation of a time series, with the goal of detecting an important change in the underlying process as soon as possible after it has occurred. In recent years there have been a growing number of papers in economics, medicine, environmental control and other areas dealing with the need of methods for surveillance. Examples are given in Frisén (1992) and Frisén (1994a). The timeliness of decisions is taken into account in the vast literature on quality control charts where simplicity is often a major concern. Also, the literature on stopping rules is relevant. For an overview, see the textbook by Wetherill and Brown (1990) the surveys by Zacks (1983) or Lai (1995) and the bibliography by Frisén (1994b).

Methods based on likelihood ratios are known to have several optimality properties. Evaluations are made of the full likelihood ratio (LR) method, which will be expressed as a certain combination of conditional likelihood ratios. In the cases studied here, the LR method has the Shiryaev optimality. Also, the Shiryaev-Roberts and the CUSUM methods are evaluated. These two methods combine conditional likelihood ratios in other ways. A comparison is also made with the Shewhart method that is a commonly used method. When control charts are used in practice, it is necessary to know several characteristics of the method. Asymptotic properties have been studied by Srivastava and Wu (1993) and Siegmund and Venkatraman (1995) and others. Here, properties for finite time of change are studied. The usual ARL^0 and ARL^1 (which are the average run lengths until an alarm under the hypothesis of no change and the hypothesis of immediate change, respectively) are used. Besides that, the probability of a false alarm, the expected delay, the probability of successful detection and the predictive value are used for evaluations. Since the methods have interesting optimality properties, the results also enlighten different criteria of optimality.

Gordon and Pollak (1995 and 1997) have suggested methods that are robust, with respect to the ARL, for different conditions of the process. Here, the robustness, with respect to optimality, for different parameter choices is studied.

In Section 1 some notations are given and there is a specification of the situation studied. In Section 2 the methods are described and their relations are demonstrated. In Section 3 there are comparisons based on a simulation study. Section 4 contains some concluding remarks.

1. NOTATIONS AND SPECIFICATIONS

The variable under surveillance is $X = \{X(t): t = 1,2,...\}$, where the observation at time t is X(t). It may be an average or some other derived statistic. The random process which determines the state of the system is denoted $\mu = \{\mu(t): t = 1,2,...\}$. As in most literature on quality control, the case of shift in the mean of Gaussian random variables from an acceptable value μ^0 (say zero) to an unacceptable value μ^1 is considered. Only one-sided procedures are considered here. It is assumed that if a change in the process occurs, the level suddenly moves to another constant level, $\mu^1>\mu^0$, and remains on this new level. That is $\mu(t) = \mu^0$ for $t= 1,...,\tau-1$ and $\mu(t) = \mu^1$ for $t= \tau, \tau+1,...$. The critical event of interest at decision time s is denoted by C(s). We want to discriminate between

$$C(s) = \{\tau \leq s\} = \{\mu(s)=\mu^1\} \text{ and } D(s) = \{\tau > s\} = \{\mu(s)=\mu^0\}.$$

Here μ^0 and μ^1 are regarded as known values and the time τ where the critical event occurs is regarded as a random variable with the density

$$\pi_t=\Pr(\tau=t)$$

and $\sum\pi_t= 1-\pi_\infty$. The intensity ν_t of a change is

$$\nu_t=\Pr(\tau=t|\tau\geq t).$$

When a specific distribution has to be specified, the geometric distribution with $\nu_t\equiv\nu$ is used.

The aim is to discriminate between the states of the system at each decision time s, s=1,2,... by the observation $X_s = \{X(s): t \leq s\}$ under the assumption that X(1) - μ(1), X(2) - μ(2),... are independent normally distributed random variables with mean zero and the same known standard deviation (say σ=1). In some calculations below, where no confusion is possible, μ^1 is denoted μ, $\mu^0=0$ and σ=1 for typographical clarity.

The values of μ and v for which a method is optimized are denoted by M and V, respectively.

Here active surveillance as defined by Frisén and de Maré (1991) is assumed. That means that the surveillance is stopped at an alarm. Thus, only one alarm is possible. A study of the properties of the first alarm at passive surveillance gives the same results. The time for the alarm is here denoted by t_A. In the literature on quality control this is usually called the run length and is denoted by RL. In the literature on probability of stopping time it is sometimes denoted by N. Its distribution under different condition contains all information about the stochastic properties of a method.

2. METHODS

2.1 The Shewhart method

Shewhart (1931) suggested that an alarm is triggered as soon as a value, which deviates too much from the target, is observed. The stopping rule is $t_A=\min(s;X(s)>G)$. This method, which is much used in quality control, has interesting relations to other methods as will be seen below. The limit G for a fixed ARL^0, is calculated by the relation: $\Pr(X(s)>G|\ \mu=\mu^0)=1/ARL^0$.

2.2 The Likelihood Ratio method

The problem of finding the method that maximizes the detection probability for a fixed false alarm probability and a fixed decision time was treated by de Maré (1980) and Frisén and de Maré (1991). The likelihood ratio (LR) method, discussed below, fulfills this criterion. Different kinds of utility functions were also discussed by Frisén and de Maré (1991). An important specification of utility is that of Girshick and Rubin (1952) and Shiryaev (1963). They treat the case of constant intensity where the gain of an alarm is a linear function of the difference $t_A-\tau$, that is the time between the alarm and the change point. The loss of a false alarm is a function of the same difference. This utility can be expressed as U= E{u(τ, t_A)}, where

$$u(\tau,t_A) = \begin{cases} l(\tau-t_A) & \text{if } \tau>t_A \\ a_1(\tau-t_A)+a_2 & \text{else.} \end{cases}$$

For the situation specified in Section 1 their solution is identical to the LR method. The "catastrophe" to be detected at decision time s is C = { $\tau \leq s$} and the alternative is D = { $\tau > s$}. The LR method has an alarm set consisting of those X_s for which the full likelihood ratio exceeds a limit. Thus

$$f_{Xs}(x_s \mid C) / f_{Xs}(x_s \mid D) = p(x_s) =$$

$$= \sum_{t=1}^{s} w_s(t) L(t) > K(s)$$

where $w_s(t)$ = Pr(τ=t)/Pr($\tau \leq s$), L(t) is the (conditional) likelihood ratio for the case when τ=t and K(s) is the limit at the decision time s.

For the case of normal distribution with C(s)={$\tau \leq s$} and D(s)={$\tau > s$}, as specified in Section 1, we have

$$p(x_s) = h(s) p_s(x_s)$$

where

$$h(s)=\frac{\exp(-(s+1)M^2/2)}{\Pr(\tau\leq s)}$$

and

$$p_s(x_s)=\sum_{k=1}^{s}\pi_k\exp\left\{\frac{1}{2}kM^2\right\}\exp\{M\sum_{u=k}^{s}x(u)\}$$

which is a nonlinear function of the observations.

In order to achieve the optimal error probabilities discussed by Frisén and de Maré (1991) an alarm should be given as soon as $p(x_s) > K(s)$.

For the case of constant intensity ν, it is possible to achieve maximum of the utility of Shiryaev, by using the limit K(s), as specified in the following definition of the LR-method.

Definition: The LR-method gives alarm as soon as

$$p(x_s)>\frac{\Pr(\tau>s)}{\Pr(\tau\leq s)}\,\frac{K}{1-K}$$

where K is a constant with respect to s. □

The alarm set can also be expressed by the posterior distribution. The relation is:

$$\Pr(\tau\leq s|X_s=x_s)>K \quad\Leftrightarrow\quad p(x_s)>\frac{\Pr(\tau>s)}{\Pr(\tau\leq s)}\,\frac{K}{1-K}$$

The Shiryaev optimality is achieved when the method is optimized for the true values, that is M=μ and V=ν. Often methods of this kind are considered Bayesian. Here the LR-method is considered to have two parameters. When they are of special interest, the method is written LR(V,M). When M is evident, the notation LR(V) is used.

Frisén and de Maré (1991) demonstrated that the alarm function of the Shewhart method is the same as that of the LR method with $C=\{\tau=s\}$. That is $p(x_s)=LR(s)$. For the normal case described in Section 1 this reduces to $p(x_s)=x(s)$.

Theorem 1: The stopping rule of the LR method tends to that of the Shewhart method when the parameter M tends to infinity.

Proof: The stopping rule with the LR method is:

$$p(X_s) > \frac{\Pr(\tau>s)}{\Pr(\tau\leq s)}\,\frac{K}{1-K}$$

$$\Leftrightarrow$$

$$\sum_{k=1}^{s} \frac{\pi_k}{\Pr(\tau\leq s)} \exp\left\{\frac{1}{2}(k-s-1)M^2\right\} \exp\left\{M\sum_{u=k}^{s} X(u)\right\} > \frac{\Pr(\tau>s)}{\Pr(\tau\leq s)}\,\frac{K}{(1-K)}$$

$$\Leftrightarrow$$

$$\exp\{(-1/2)M^2\}\exp\{MX(s)\}$$

$$> \frac{\Pr(\tau>s)}{\pi_s}\,\frac{K}{(1-K)} - \sum_{k=1}^{s-1} \frac{\pi_k}{\pi_s} \exp\left\{\frac{1}{2}(k-s-1)M^2\right\} \exp\left\{M\sum_{u=k}^{s} X(u)\right\}$$

$$\Leftrightarrow$$

$$\exp\{MX(s)\}$$

$$> \exp\{M^2/2\}\frac{\Pr(\tau>s)}{\pi_s}\,\frac{K}{(1-K)} - \sum_{k=1}^{s-1} \frac{\pi_k}{\pi_s} \exp\left\{\frac{1}{2}(k-s)M^2\right\} \exp\left\{M\sum_{u=k}^{s} X(u)\right\}$$

$$\Leftrightarrow$$

$$\exp\{MX(s)\} > \exp\left\{\frac{1}{2}M^2\right\}\left\{\frac{K\Pr(\tau>s)}{(1-K)\pi_s} + o\left(\frac{1}{M}\right)\right\}$$

$$\Leftrightarrow$$

$$X(s) > \frac{M}{2} + \frac{1}{M}\ln\left\{\frac{K\,(1-V)}{(1-K)V} + o\left(\frac{1}{M}\right)\right\}$$

which tends to the stopping rule of the Shewhart method for large values of M, since the dependency on s, of the alarm limit for X(s), disappears.□

2.3 Shiryaev-Roberts method

Shiryaev (1963) and Roberts (1966) suggested that an alarm is triggered at the first time s, for which

$$\sum_{t=1}^{s} L(t) > K$$

where K is a constant.

This method is the limit of the LR method

$$\sum_{t=1}^{s} w_s(t)L(t) > K(s)$$

when V tends to zero since both the weights $w_s(t)$ and the limit K(s) tend to constants. Roberts (1966) motivated the method by the conjecture that the intensity parameter ν has very little influence on the LR method (for ν less than 0.2) and thus the weights which depend on ν can be omitted. Pollak (1985) proved that the Shiryaev-Roberts method is asymptotically Bayes risk efficient as $V \rightarrow 0$. Pollak (1985) suggested a modification of the Shiryaev-Roberts procedure by a stochastic starting value which might be considered as reflecting the prior belief regarding the likelihood of a change when surveillance is initiated. Pollak (1985) and Yakir (1997) have proven different optimality properties such as asymptotically (as $K \rightarrow \infty$) minimax properties for this modification. Here however, the original version is studied.

Theorem 2: The stopping rule of the Shiryaev-Roberts method tends to that of the Shewhart method when M tends to infinity.

Proof: In analogy with the proof of Theorem 1.

2.4 CUSUM

The cumulative sum

$$C_t = \sum_{i=1}^{t} (X(i) - \mu^0)$$

is used in several CUSUM- variants. The most commonly advocated variant gives an alarm for the first t for which $C_s - C_{s-i} > h + ki$ (for some i=1,2,... s), where $C_0=0$ and h and k are chosen constants (Page 1954).

Sometimes the CUSUM test is defined by likelihood ratios (e.g. Siegmund 1985 and Park and Kim 1990). An alarm is given as soon as

$$\max(L(t)\ ;\ t=1,2,....s) > K$$

where L(t), as defined above, is the likelihood ratio corresponding to a change at time t, and K is a constant. The last definition reduces to the first definition above in the case specified in Section 1 with normal distribution if k=M/2. In this simulation study the parameter k=M/2 is used as this often is considered to be optimal.

Theorem 3: When the value of M for which the CUSUM method is optimized tends to infinity, then the properties tend to be identical to those of the Shewhart method.

Proof:

$$M \to \infty \ \Rightarrow\ k = \frac{M}{2} \to \infty \ \Rightarrow$$

$$\Pr(C_s - C_{s-i} > h + ki;\ i=1,...s) \to \Pr(C_s - C_{s-i} > ki;\ i=1,...s) = \Pr(X(s) > k)$$

$$\text{since } t_A \geq s \Rightarrow X(i) \leq k \ \ i=1,...s-1;\ \Box$$

3. RESULTS OF SIMULATIONS

For the Shewhart method analytical calculations were made. For the other methods simulation of 10 000 000 replicates were made for each of the (more than 1000) situations studied. The large number of replicates was necessary in order to make all results and figures reliable. The confidence intervals for all results are now small compared with the thickness of the lines of the figures. The first pont of most curves is exactly calculated. For $ARL^0 = 11$ change points 1, 2,..., 40 and many variants of the methods were studied. The short ARL^0 used in most of the studies was chosen to make the computer time, necessary for the study, reasonable. For $ARL^0 = 100$ fewer cases were studied. The generality of the results for different values of the ARL^0 is discussed in Section 3.4. The limits of the stopping rules were determined with an exactness which made the deviation between the intended and estimated ARL^0 less than (and for most cases much less than) 0.1 % of the intended value. The number of replicates makes it possible to neglect the sampling error for this determination.

3.1 Probability of a false alarm

In all simulations the parameters of the different methods were chosen so that ARL^0 was equal for all methods (equal to 11 in most comparisons and 100 in some) to make them comparable. To fix the value of ARL^0 is in accordance with most comparative studies in this area. However, as will be seen, it is not the only alternative.

Equal values of ARL^0 do not imply that the run length distributions are identical when $\mu \equiv \mu^0$. The distributions can have different shapes. This is demonstrated in Figure 1. Here the probabilities of an alarm at a specific time point, when no change has occurred, are given for some different methods. The skewness of the distributions is less pronounced for the LR method with a great intensity parameter. For $ARL^0 = 11$, the distributions for the Shewhart and the CUSUM methods are very similar except at the first point.

A summarizing measure of the false alarm distribution is the total probability of a false alarm,

$$\Pr(t_A<\tau)=\sum_{t=1}^{\infty}\Pr(\tau=t)\Pr(t_A<t|\tau=t)$$

It is illustrated as a function of ν in Figure 2 for the case where the probability of a change has a geometric distribution with the intensity ν. The first factor in the sum does not depend on the method but only on the true intensity ν. The second factor depends only on the run length distribution when no change has occurred. Since the values of the ARL^0 are equal for all methods only the different shapes and not their locations will influence the false alarm probability. Thus, only the size of the differences in Figure 2 can be expected. Although the number of replicates was as large as 10 000 000, this is not enough for reliable values for the situation with very small ν. However, the limiting value, when ν tends to zero, is easily calculated to be one. This value, with linear interpolation, was used for the extremely small values of ν. Asymptotic values , as ARL^0 increases, for the Shiryaev-Roberts method as given by the relation $ARL^0 = (1-\Pr(t_A<\tau))/\nu\Pr(t_A<\tau)$ in Kolmogorov et al. (1990) are not good enough for ARL^0 =11 or 100.

3.2 Delay of an alarm

As was seen above, the correspondence to the level of significance in an ordinary test is not a value but the run length distribution, namely the distribution of t_A , the time of an alarm. For the power the correspondence is still more complicated. To describe the ability of detecting a change we need a set of run length distributions for different times of change.

The case where the change occurred before the surveillance started (τ=1), is illustrated in Figure 3 for some methods. The effect of values of the parameter M different from the true μ can be seen. This illustrates the robustness of this parameter choice. The speed of the limiting behaviour, which was proven above, is illustrated.

The average run length under the alternative hypothesis, $E(t_A|\ \mu \equiv \mu^1) = E(t_A|\tau=1) = ARL^1$, is the mean number of decisions that must be taken to detect a change (that occurred at the same time as the inspection started). The part of the definition in the parenthesis is seldom spelled out but seems to be generally used in the literature on quality control. The values of ARL^1 for the case of $ARL^0=11$ are given in Figure 4. Here the convergence with increasing M to the properties of the Shewhart method is clearly seen. For $ARL^0=100$ only few values are available but it is clear that the convergence is much slower.

Since the case $\tau=1$ is not the only case of interest the expected delay is calculated also for other values of τ. Some kind of summarizing measure is useful. The expected delay

$$E(t_A-\tau|t_A\geq\tau,\ \tau=t)$$

is given in Figure 5 for different values of τ and M. For $\tau=1$ the values of this function equal the values of $(ARL^1 - 1)$. The differences in shapes of these curves demonstrate the need for other measures than the conventional ARL. The case of $ARL^0=11$ and M=1 is a good example of this. Although the Shiryaev-Roberts method has worse ARL^1, and thus worse delay for a change at $\tau=1$, than the Shewhart method, it is better for all other times of change. Another example is that the CUSUM and the Shewhart methods are very much alike for $\tau=1$, for $ARL^0=11$, but the CUSUM is better for all other change points.The convergence with increasing M to the expected delay of the Shewhart method is much slower for the larger ARL^0. This is in accordance with the results for ARL^1.

The expectation of the delay also with the respect to the distribution of τ is:

$$E(t_A-\tau|t_A\geq\tau)$$

This function is given as a function of ν in Figure 6, for the case when the distribution of τ is geometrical with the intensity ν. The estimates by the simulations for very small ν are not reliable in spite of the large number of replicates. However, then ν tends to one, the expected delay tends to ARL^1-1 and this value and a linear interpolation was used for the extremely small values of ν.

In some applications, such as medical intensive care, there is a limited time available for rescuing actions. Then, the expected value of the difference τ-t_A is not of main interest. Instead, the probability that the difference does not exceed a fixed limit is used. The fixed limit, say d, is the time available for successful detection.

The probability of successful detection,

$$\mathbf{PSD(t,d)}=\Pr(t_A-\tau<d\,|\,t_A\geq\tau,\ \tau=t).$$

was suggested by Frisén (1992) as a measure of the performance. It is illustrated in Figure 7.

3.3 Predicted value

The predictive value $PV(t) = \Pr(\tau \leq t \,|t_A=t)$ has been used as a criterion of evaluation by Frisén (1992), Frisén and Åkermo (1993) Åkermo (1994) and Frisén and Cassel (1994). It is illustrated in Figure 8. The price for a high probability of fast detection of a change in the beginning of the surveillance (as was demonstrated in Figure 7) for the CUSUM and the Shewhart method is that the early alarms are not reliable. The predicted value of an alarm depends on the intensity ν of the process. As will be seen in the next section, different values of ν have to be used for different values of ARL^0 to make comparisons easier. If the same value had been used, the predicted values for ARL^0=11 had been much lower than those for ARL^0=100.

3.4 Effect of different values of the ARL^0

Most of the simulations were made for a very low value of ARL^0 to make the computer time reasonable. This corresponds in a way to making observations more seldom. For example, instead of making them each minute they may be made each hour. A low ARL^0 thus makes continuous time approximations less reliable. Pollak and Siegmund (1985) derive some results that seem impossible to compute in discrete time. Some recent publications on change-point detection in continuous time are Chu et al. (1996) and Beibel (1997). The change in time scale may also have other effects which will now be discussed.

The ordering between the methods with respect to early alarms is the same for both the values of ARL^0 used in the simulations. The Shewhart method has the largest probability for an early alarm. It is followed in order by the CUSUM, the Shiryaev-Roberts and the LR-methods with increasing V. However for the small ARL^0=11 the Shewhart method was very similar to the CUSUM method and the important difference was between these two methods and the others. For ARL^0=100 this is no longer the case. An explanation for this is that the relative information loss of using only the last observation is much larger for the larger ARL^0 which corresponds to frequent observations. For the larger ARL^0, the major difference is between the Shewhart method and the others.

The results on probability of false alarm, expected delay and the predicted value which are direct functions of the true intensity ν will be affected by different ARL^0 since both are dependent on the time scale.

The results on robustness for the choice of the intensity V are heavily dependent on the size of the ARL^0 since ν, V and ARL^0 all relate to the time scale. One possible way to relate them is by a fixed value of $\Pr(\tau \geq ARL^0)$. This relation will not make the situations perfectly comparable because of the discreteness involved. However, one effect of the time scale will be considered. Let ν_a and ν_b be the corresponding intensities for situations with ARL^0 equal to a and b respectively. Then

$$\Pr(\tau > a \mid \nu = \nu_a) = \Pr(\tau > b \mid \nu = \nu_b))$$
$$\Leftrightarrow$$
$$(1-\nu_a)^a = (1-\nu_b)^b$$

For example, by this relation the value ν_{11}=0.6 corresponds to ν_{100} = 0.1. Also other relations were studied. Examples are fixed false alarm probability for the Shewhart method, fixed ratio $w(ARL^0)/w(1)$ between the weights in the LR-method, and fixed ratio $K(ARL^0)/K(1)$ between the limits for alarm, for s=ARL^0 and s=1, by the LR-method. All these gave approximately the same numerical results. However, for the similarities between the Shiryaev-Roberts and the LR method it is not only the conditions for time points far apart which influence, but also the similarities between adjacent weights. This works in the other direction and the combined effects observed in the figures are that ν_{11}=0.2 (or even a little less) corresponds better to $\nu_{100} = 0.1$.

The influence of ARL^0 on the robustness for the choice of V is evident since both are connected to the time scale. Even if less pronounced there is also a connection between the time scale and M since the optimal method for large M is to allocate much alarm probability for early time points and vice versa. However, the theorems on the similarities between methods, when the parameters are determined for large values M, does not depend on the size of ARL^0. As was discussed above, the difference between the Shewhart method and the others (with M=μ) is more pronounced for large ARL^0 . Thus M is more important for large ARL^0 since the behaviour for large M and M=μ differs more. This is illustrated by the figures.

4. DISCUSSION

Sometimes the LR method is considered a Bayesian method while the Shiryaev-Roberts method is considered a frequentistic one. Here, however no assumptions are made which automatically violates the frequentistic framework. No specially Bayesian assumptions are necessary for the LR method. The change point τ can be regarded as an ordinary stochastic variable. It is not necessary to interpret the intensity parameter V as a Bayesian assumption of the process. The properties of the method will be better if the parameter is not far from the actual intensity but the method can be considered anyhow.

The number of parameters for the studied methods differs. The Shewhart method has no parameters which can be used to optimize the method. The only adjustment that can be made is to set the limit for an alarm to make the average false alarm properties (here ARL^0) comparable with other methods. The CUSUM and the Shiryaev-Roberts methods have one parameter to optimize for the size of the shift μ but none for the intensity ν. First, the effect of the optimizing for μ will be discussed, then the effect of the optimizing for ν, and finally the general properties will be considered.

The case where μ is unknown was examined by Srivastava and Wu (1993). They studied the effect of different true μ for a fixed assumed M. Here, the properties for different values of M for a fixed μ are studied to examine the robustness to the choice of parameter value M. The theorems and the figures demonstrate that choice of a large value of M makes the properties of the methods more alike. That means that a wrong specification to a too large value is not very dramatic as concerns the choice of method. All methods will have very skew distributions with high probabilities of early alarms. On the contrary, when the methods are optimized for small values, the skewness of the run length distribution is shifted and very early alarms are more rare. The skewness of the distributions becomes less pronounced compared with that of the Shewhart method. The differences in performance between the Shewhart method and the others is more pronounced for the large ARL^0 than for the smaller one because of the relatively larger information loss. Thus it is not surprising that the convergence with increasing M to the properties of the Shewhart method is slower for the large ARL^0.

The less skewness for smaller values of M corresponds to a smaller probability of a false alarm $Pr(t_A<\tau)$ for a fixed ARL^0. For late changes the power of detection gets better as is reflected by the expected delay and the probability of successful detection. The total expected delay, for the optimal values of the parameters, is optimal for the LR method for a fixed probability of false alarm. Now the ARL^0 is fixed instead. The balance between the false alarms and the power, as reflected by the predicted value, is better for the Shiryaev-Roberts and LR methods for early alarms.

The LR method is the only method here which has a parameter to optimize for the intensity ν. That the Shiryaev-Roberts method is the limit of the LR method when the intensity ν tends to zero is seen by the formulas and illustrated by the simulated results. The smaller the intensity parameter, the smaller the difference between the LR method and the Shiryaev-Roberts method. Simulations were made also for ν=0.001 and ν=0.01 but these results are not included in the figures since the differences to those of the Shiryaev-Roberts method are less than the line width. The Shiryaev-Roberts method is a good approximation of the optimal LR method and thus approximately optimal, for small values of ν.

The figures demonstrate that the LR method is robust against wrong specification of the intensity parameter V for the cases studied. In comparison with the difference to the Shewhart method, the choice of the intensity parameter of the LR method has very little influence on the performance. The results here confirm the conjecture by Roberts (1966) about the robustness of the LR method with respect to V.

The optimal method when the intensity is large should intuitively have a large probability of early alarms. Some present results might seem to be contradicting this. These results will now be discussed.

The shapes (see Figures 1 and 3) of the distributions of the alarm time t_A might seem surprising. The probability of a very early alarm by the LR method with a low intensity parameter is greater than for a large value of the parameter. However, for a low intensity the probability of a late change is great and thus a thick tail of the distribution of t_A is appropriate. As the expected value ARL^0 is fixed, the only possibility is a high probability of early alarms. This also causes the differences in false alarm probabilities in Figure 2.

The large expected delay (see Figures 5 and 6) for the LR method with a large intensity parameter ν might seem surprising. For a fixed cost of a false alarm the delay will be less for greater values of the intensity parameter ν. However, now we have a fixed ARL^0, and thus a greater value of this cost, which implies that the delay is increased to make the locations of the distributions of t_A equal. The difference between the distributions is the shape, and not the location, and this difference causes the expected delay to be greater for the greater values of the intensity parameter.

The predicted value of an alarm is given in Figure 8. This reflects the trust one should have in an alarm. A constant predicted value makes the interpretation of an alarm easier since the same kind of action is appropriate both for early and late alarms. The Shiryaev-Roberts and the LR-methods have relatively constant predicted values when $M=\mu$. This advantage is however diminished if a wrong specification of μ is used.

The conclusion by Mevorach and Pollak (1991), that the expected delay is similar for the Shiryaev-Roberts and the CUSUM methods also for small values of ARL^0, can only be confirmed for large values of M. They studied the cases M=5 and M=7 for $\mu=1$. In the present study the conclusion is that there is similarity for the small $ARL^0=11$ only for large values of M. Then, all methods converge to the Shewhart method and become similar. Also, they studied a situation where the time of change, τ, does not have any influence. When the effect of τ is considered the CUSUM in many aspects was very similar to the Shewhart method. Srivastava and Wu (1993) made the conclusion that for small ARL^0, Shiryaev-Roberts methods is clearly better than the CUSUM method.

Most theoretical comparisons are made for the case where ARL^0 tends to infinity. Even though $ARL^0=100$ is not large enough for the results about measures of performance to be approximated by those results some general conclusions agree. The conclusion by Pollak (1985), that the CUSUM and the Shiryaev-Roberts methods are similar for large ARL^0, is confirmed. The relative information loss with the Shewhart method, that only takes advantage of the last observation, makes the properties of the Shewhart method more different from those of the others. Even though the CUSUM method still resemblances the Shewhart method in the tendency of a power-spending to early alarms, it is now much more alike the others than the Shewhart method.

At large, the properties differ between the LR and the Shiryaev-Roberts methods on one hand and the Shewhart method on the other, for the situation studied in the simulations. This difference is more pronounced for the large ARL^0 and for the cases where a change has occurred. This is not surprising since the loss of information by using only the last observation is worst for these cases. The main difference is the "error-spending" on early or late alarms. When the common practice in quality control, to use ARL^0 and ARL^1 only to judge methods, this fact is not demonstrated. Neither the effect of the skewness nor the effect of the value of τ is seen by those measures. For Shewhart early (but often false) alarms are more frequent. The drawback with too early alarms is reflected by the predicted value. The early alarms are of less value since the predicted value is very low.

REFERENCES

Chu, C. Stinchcombe M. & White H. (1996) Monitoring structural change. *Econometrica*, 1045-1065.

Beibel, M. (1997) Sequential change-point detection in continuous time when the post-change drift is unknown. *Bernoulli*, 4, 457-478.

Frisén, M. (1992), Evaluations of methods for statistical surveillance, *Statistics in Medicine*, 11, 1489-1502.

Frisén, M. (1994a) Statistical Surveillance of Business Cycles. *Research report*, 1994:1, Department of Statistics, Göteborg University.

Frisén, M. (1994b) A classified bibliography on statistical surveillance. *Research report*, 1994, Department of Statistics, Göteborg University.

Frisén, M. and Cassel, C. (1994) Visual evaluations of statistical surveillance. *Research report*, 1994:3, Department of Statistics, Göteborg University.

Frisén, M. and de Maré, J. (1991), Optimal surveillance, *Biometrika*, 78, 271-280.

Frisén, M. and Åkermo, G. (1993), Comparison between two methods of surveillance: exponentially weighted moving average vs CUSUM, *Research report*, 1993:1, Department of Statistics, Göteborg University.

Girshick, M. A. and Rubin, H. (1952), A Bayes approach to a quality control model, *Ann. Math. Statist.*, 23, 114-125.

Gordon, L. & Pollak, M. (1995) A Robust surveillance scheme for stochastically ordered alternatives. *Annals of Statistics*, 23, 1350-1375.

Gordon, L. & Pollak, M. (1997) Average run length to false alarm for surveillance schemes designed with partially specified pre-change distribution. *Annals of Statistics*, 25, 1284-1310.

Maré, J. de (1980), Optimal prediction of catastrophes with application to Gaussian processes, *Ann. Prob.* 8, 841-850.

Kolmogorov, A. N., Prokhorov, Y. V. & Shiryaev, A. N. (1990) Probabilistic-statistical methods of detecting spontaneously occurring effects. *Proceedings of the Steklov Institute of Mathematics* 1, 1-21.

Lai, T.L. (1995), Sequential changepoint detection in quality control and dynamic systems, *J.R. Statist. Soc. B* 57, 613-658.

Mevorach, Y. and Pollak, M. (1991), A small sample size comparison of the CUSUM and Shiryaev-Roberts approaches to changepoint detection, *American J. of Mathematical and Management Sciences*, 11, 277-298.

Page, E. S. (1954), Continuous inspection schemes, *Biometrika*, 41, 100-114.

Park, C. S. and Kim, B. C. (1990) A CUSUM chart based on log probability ratio statistic, *J. Korean Statistical Society*, 19, 160-170.

Pollak, M. (1985) Optimal stopping times for detecting changes in distributions. *Ann. Statist.* 13, 206-227.

Pollak, M. & Siegmund, D. (1995) A diffusion process and its applications to detecting a change in the drift of Brownian motion. *Biometrika*, 72, 267-280.

Roberts S. W. (1966) A comparison of some control chart procedures. *Technometrics*, 8, 411-30.

Shewhart, W. A. (1931) *Economic Control of Quality Control*. Reinhold Company, Princeton N.J.

Shiryaev, A. N. (1963), On optimum methods in quickest detection problems, *Theory Probab. Appl.*, 8, 22-46.

Siegmund, D. (1985), *Sequential analysis. Tests and confidence intervals*, Springer.

Siegmund, D. & Venkatraman, E. S. (1995) Using the generalized likelihood ratio statistic for sequential detection of a change-point. *Annals of Statistics*, 23, 255-271.

Srivastava, M.S. and Wu, Y (1993) Comparison of EWMA, CUSUM and Shiryaev-Roberts procedures for detecting a shift in the mean. *Ann. Statist.*, **21**, 645-670.

Wetherill, G.B. and Brown, D.W. (1990), *Statistical process control*, London: Chapman and Hall.

Wessman, P. (1996) Some principles for surveillance adopted for multivariate processes with a common change point. *Research report*, 1996:2, Department of Statistics, Göteborg University.

Yakir, B. (1997) A note on optimal detection of a change in distribution. *Ann. Statist.*, **25**, 2117- 2126.

Zacks, S. (1983) Survey of classical and Bayesian approaches to the change-point problem: Fixed sample and sequential procedures of testing and estimation. in *Recent advances in statistics*, 245-269.

Åkermo, G. (1994), Constant predictive value of an alarm, *Research report*, 1994:6, Department of Statistics, Göteborg University.

DEPARTMENT OF STATISTICS
GÖTEBORG UNIVERSITY
P.O. BOX 660
SE-40530 GÖTEBORG
SWEDEN

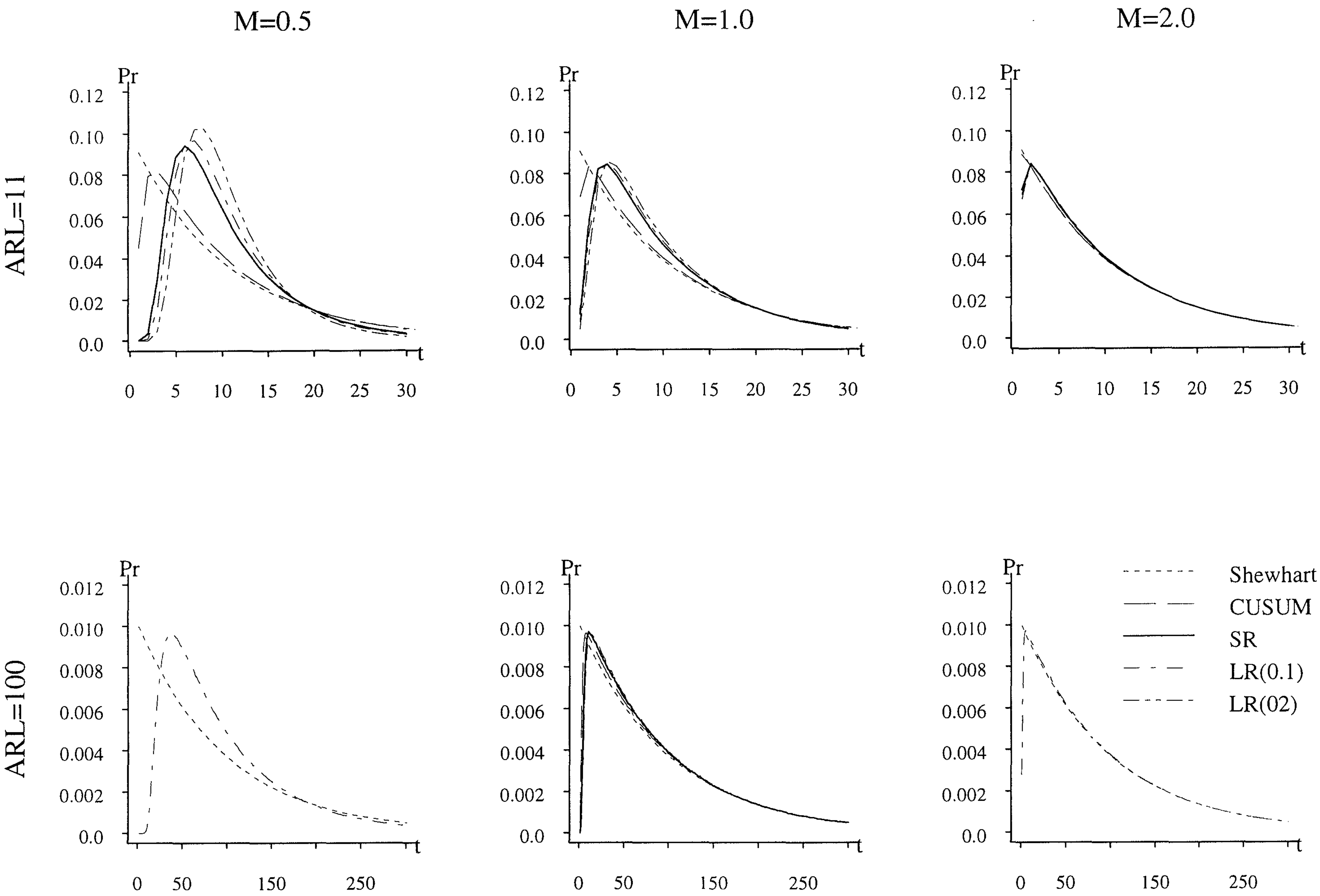

Figure 1. The density of the time of alarm, $Pr(t_A=t|\mu_i\equiv\mu^0=0)$. This is the probability of an alarm at time t, when no change has occurred. The expectations, ARL^0, of the distributions are all equal to 11 in the upper row and equal to 100 in the bottom row. In the columns M=0.5, 1 and 2 respectively. The Shiryaev-Roberts method is denoted SR.

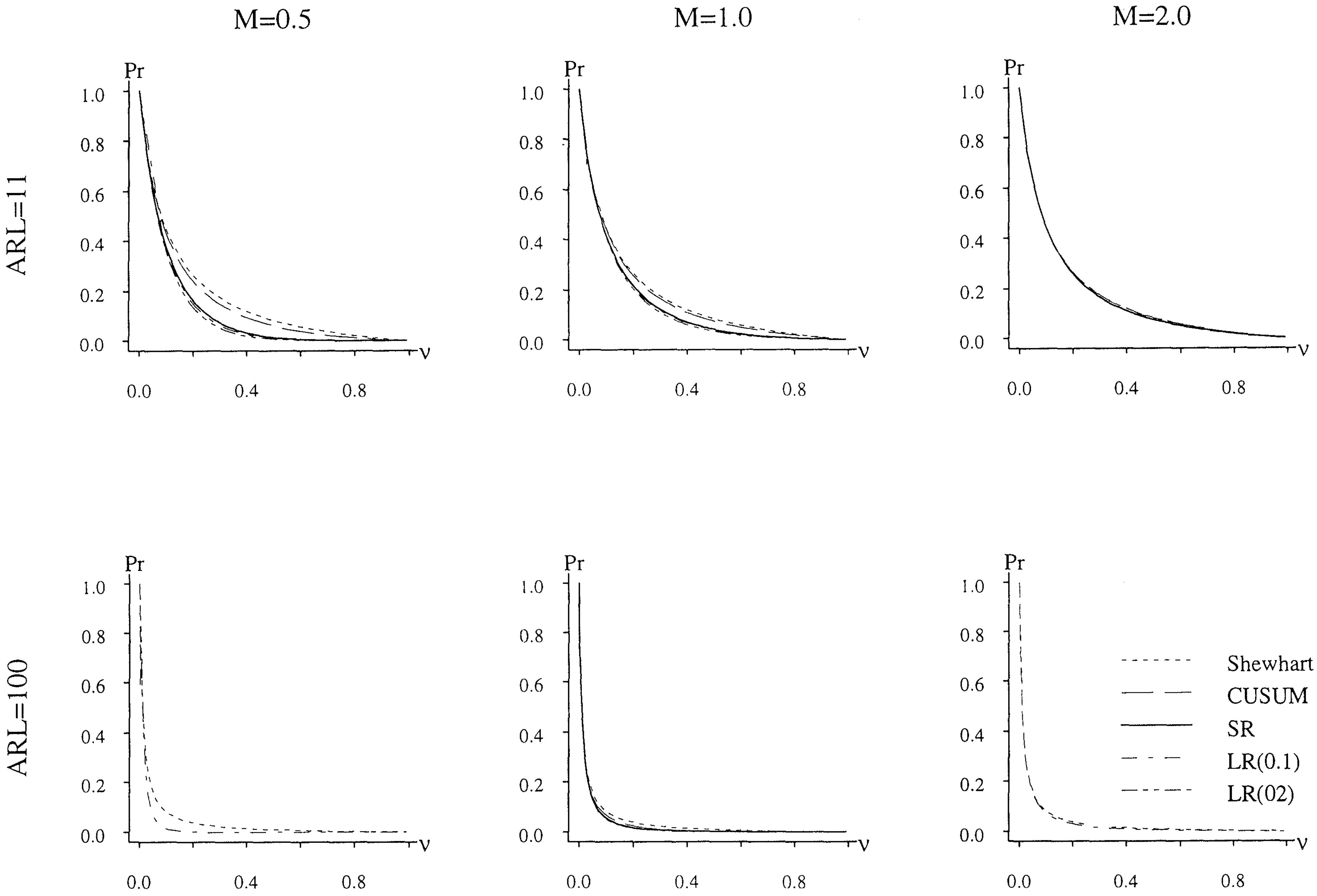

Figure 2. The expected probability of a false alarm $Pr(\tau>t_A)$ when the distribution of τ is geometrical with intensity ν.

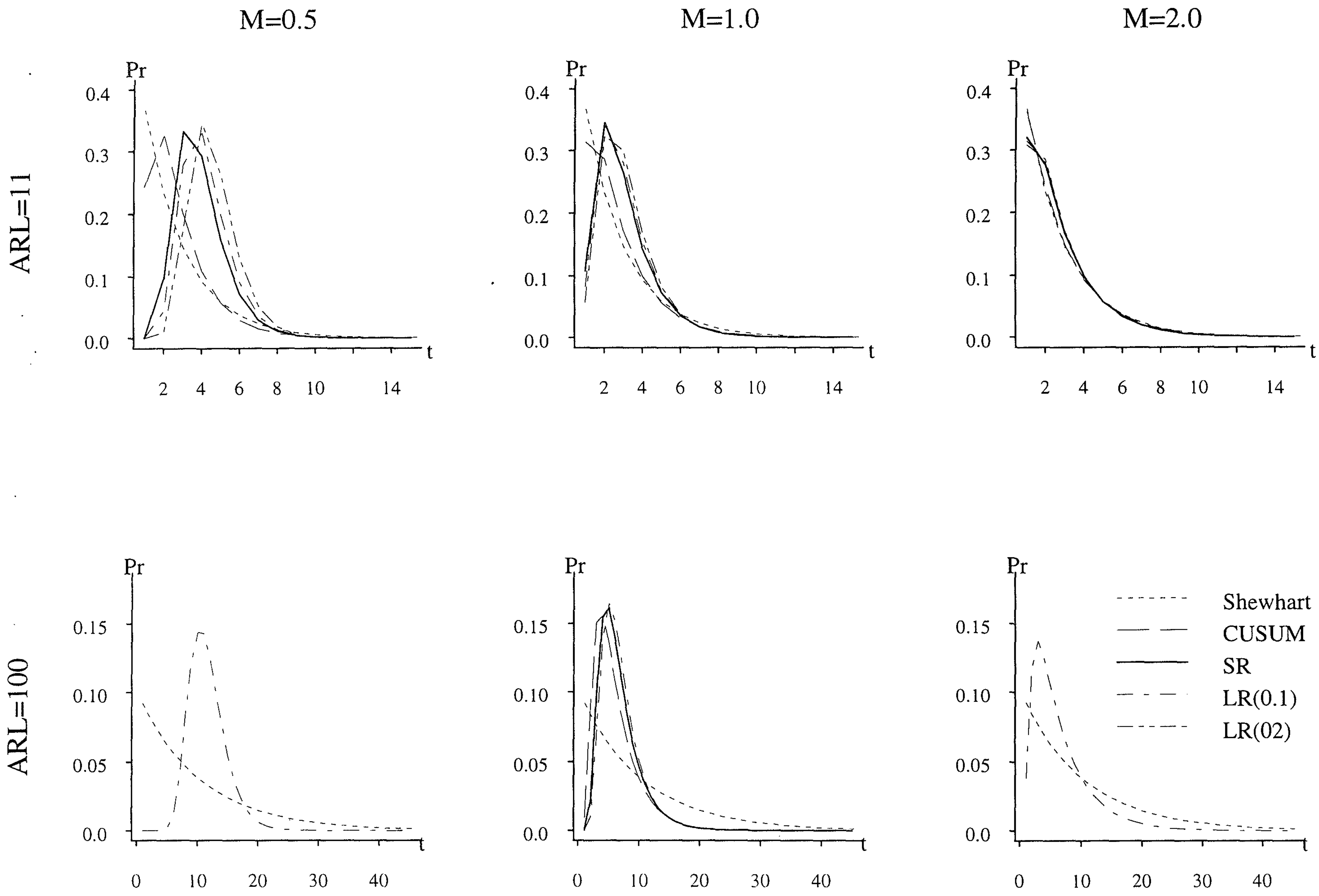

Figure 3. The density of the time of an alarm, when the change occurred before the surveillance started, $Pr(t_A{=}t|\ \mu_i \equiv \mu^1 = 1)$.

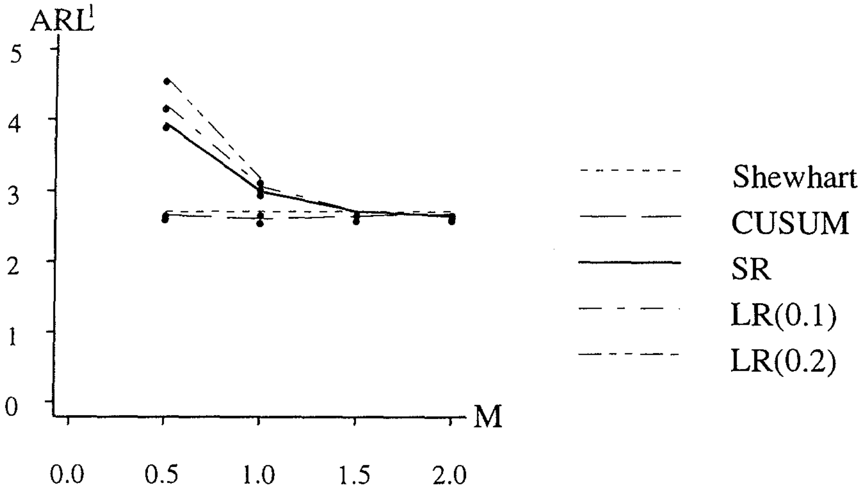

Figure 4. The ARL^1 as a function of M for $ARL^0 = 11$.

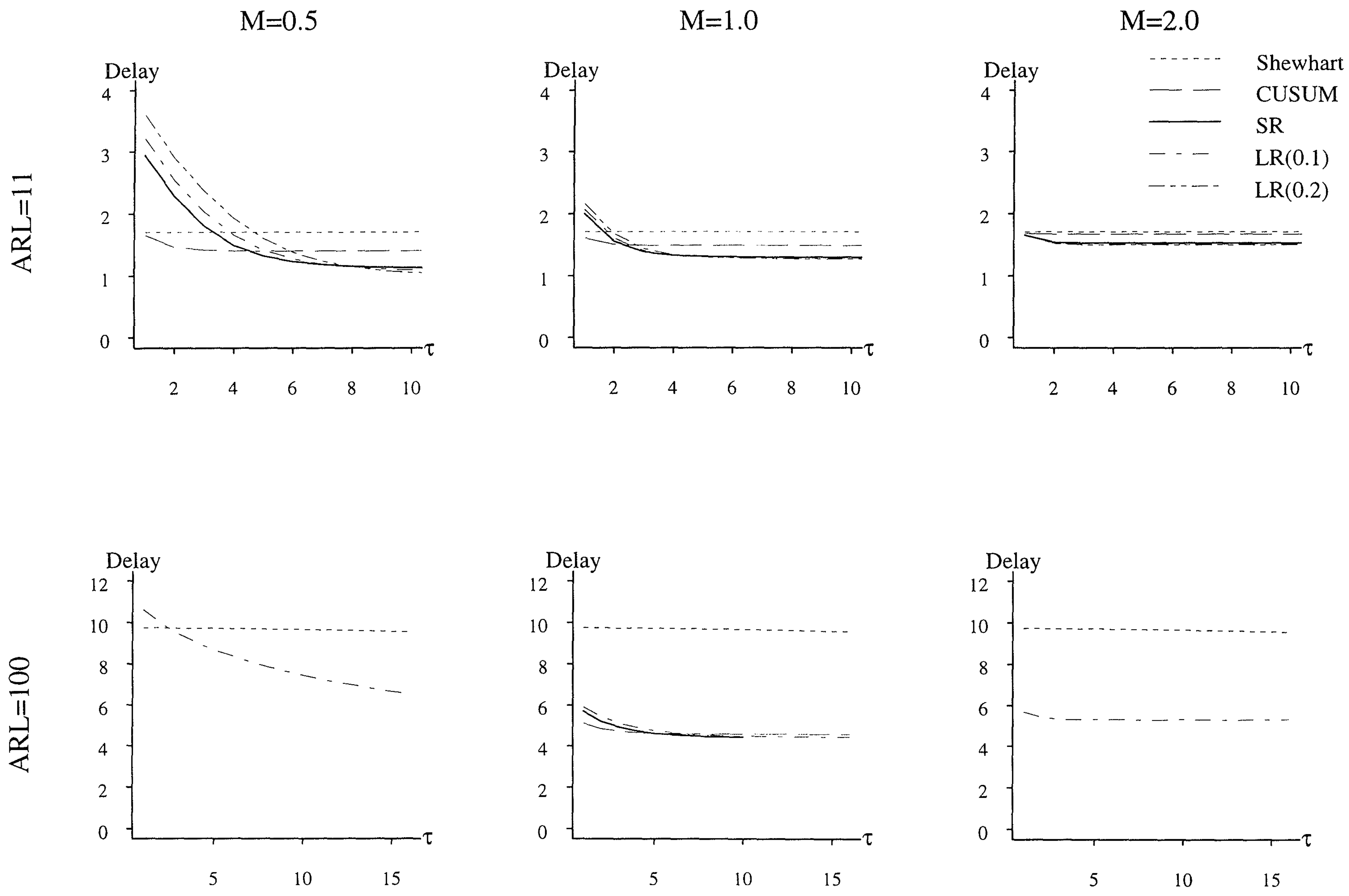

Figure 5. The expected delay after a change at time τ=t. That is: $E(t_A - \tau | t_A \geq \tau, \ \tau)$.

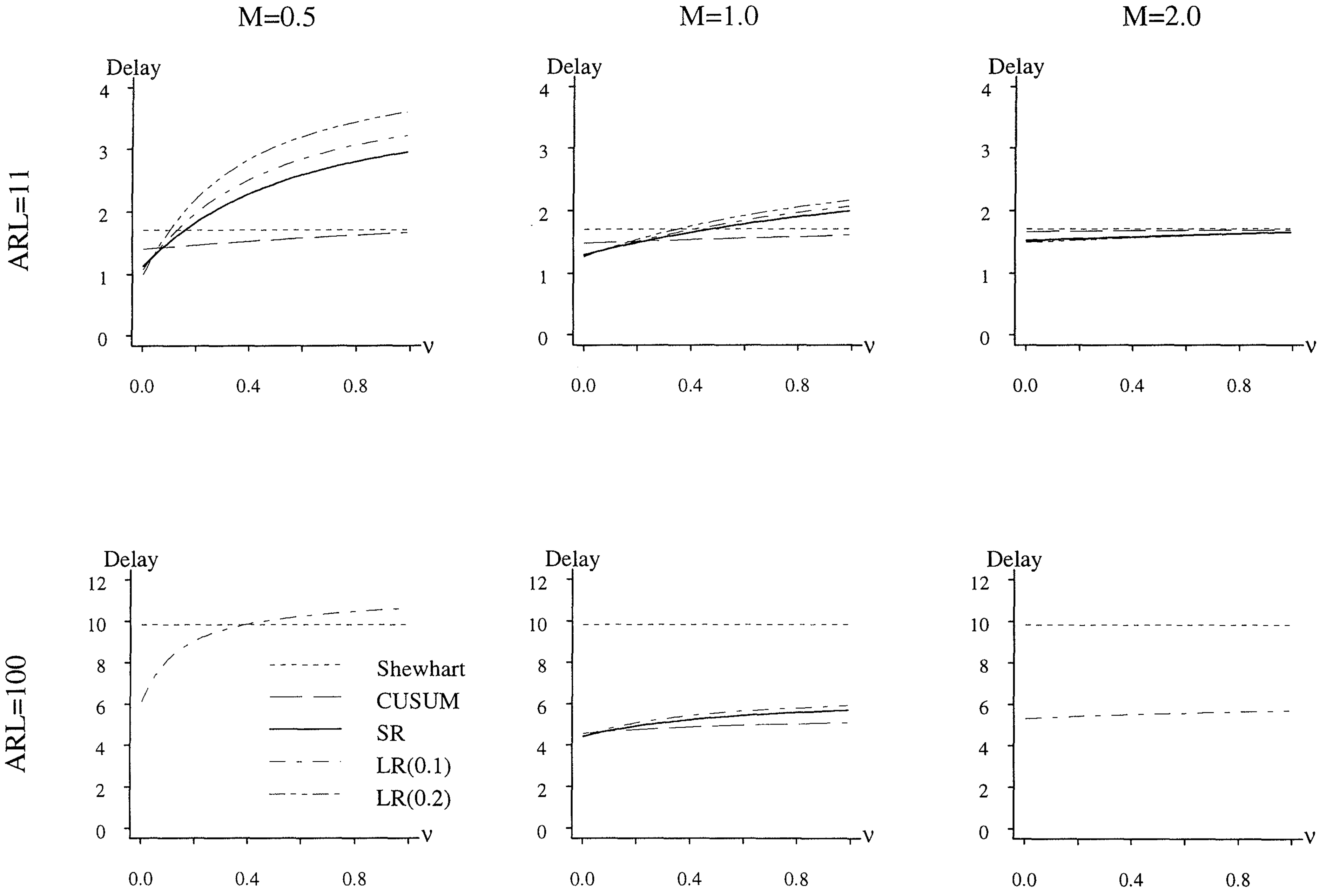

Figure 6. The expected delay when τ has a geometric distribution with intensity ν. That is: $E(t_A - \tau | t_A \geq \tau)$

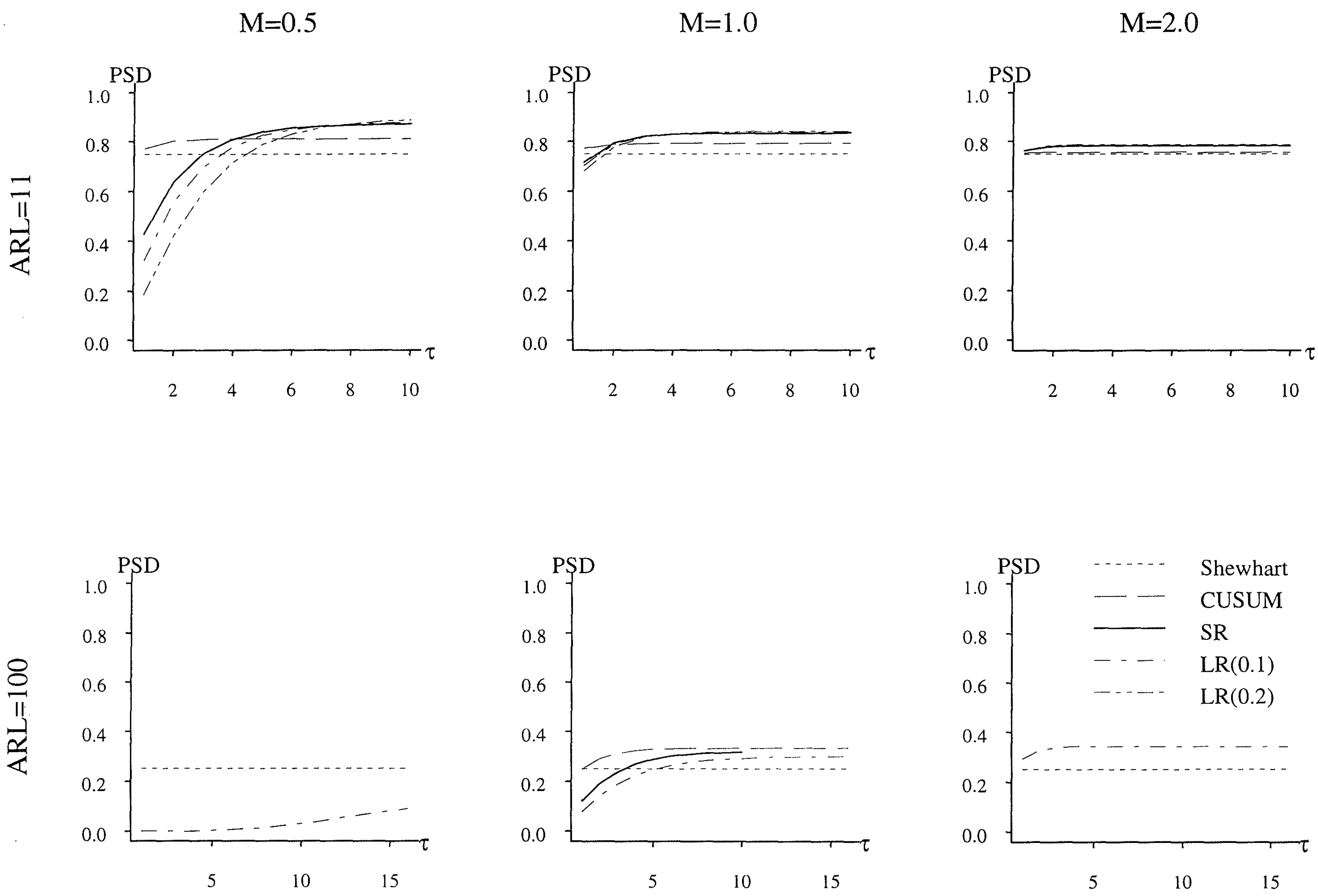

Figure 7. The probability of successful detection, $PSD(\tau,d)=Pr(t_A-\tau<d \mid t_A\geq\tau,\tau)$, for d=2.

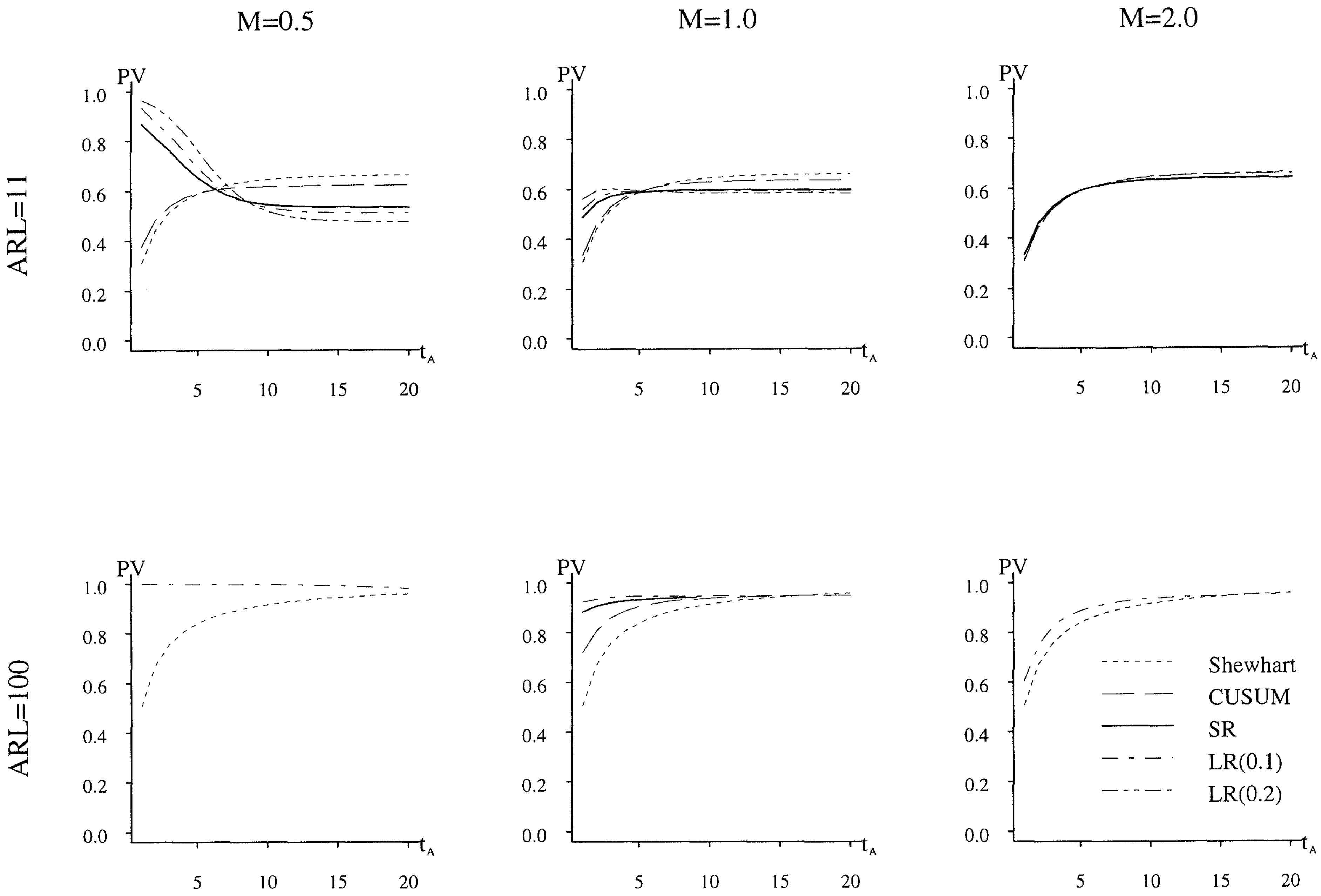

Figure 8. The predictive value $PV(t_A) = Pr(\tau \leq t \mid t_A)$ is given for the case when τ has a geometric distribution. The intensity ν is 0.2 in the upper row there $ARL^0 = 11$ and 0.1 in the bottom row there $ARL^0 = 100$.

Research Report

1995:1	Arnkelsdottir, H	Surveillance of rare events. On evaluations of the sets method.
1995:2	Sveréus, A	Detection of gradual changes. Statistical methods in post marketing surveillance.
1995:3	Ekman, C	On second order surfaces estimation and rotatability.
1996:1	Ekman, A	Sequential analysis of simple hypotheses when using play-the-winner allocation.
1996:2	Wessman, P	Some principles for surveillance adopted for multivariate processes with a common change point.
1996:3	Frisén, M. & Wessman, P	Evaluations of likelihood ratio methods for surveillance.
1996:4	Wessman, P.	Evaluation of univariate surveillance procedures for some multivariate problems.
1996:5	Särkkä, A.	Outlying observations and their influence on maximum pseudo-likelihood estimates of Gibbs point processes.
1997:1	Ekman, A.	Sequential probability ratio tests when using randomized play-the-winner allocation.
1997:2	Pettersson, M.	Monitoring a Freshwater Fishpopulation: Statistical Surveillance of Biodiversity.
1997:3	Jonsson, R.	Screening-related prevalence and incidence for non-recurrent diseases.
1997:4	Johnsson, T.	A stepwise regression procedure applied to an international production study - a multiple inference solution.
1998:1	Särkkä, A. & Högmander, H.:	Multiple spatial point patterns with hierarchical interactions.