Research Report
Department of Statistics
Göteborg University
Sweden

The surveillance of several processes with different change points

Peter Wessman

Research Report 1999:2
ISSN 0349-8034

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

The surveillance of several processes with different change points.

Peter Wessman
Department of Statistics, Göteborg University.
Box 660, S 405 30, Göteborg, Sweden.

1998

Abstract

A statistical surveillance situation which involves the simultaneous surveillance of several processes is treated. Some recently suggested multivariate methods are discussed together with a new method based on the likelihood ratio. The emphasis in the discussion is put on different ways to combine information from each time point. The methods treated represent different approaches in this aspect to the construction of multivariate surveillance methods. Shewhart type methods are used to handle the information over time. Comparisons of these methods are made when two processes, which are observed through bivariate normal variables, are surveilled for sudden shifts in the means with known and constant covariance structure. Also, the effects of different change points for the variables are considered. Generalisations to the multivariate case of the ARL and the probability of a successful detection are suggested. The main difference in performance between the compared methods is shown to be between methods based on the marginal and joint distributions of the variables. It is also shown how the choice of method depends on both on the correlation between the variables and when the time when a second change point can be expected.

Acknowledgement 1 *I am grateful to Prof. M. Frisén for her suggestions and encouragement at all stages in the preparation of this work.*

1 Introduction

Statistical surveillance considers a problem that appears in most fields of application; a random process $\theta(t)$ in time must be monitored to detect an important change as quickly as possible and with a minimum of false alarms.

A variable $X(t)$ with some distribution $F_{\theta(t)}$ dependent on this process is therefore observed sequentially over time. After each new observation a decision based on these observations is made on whether or not a change has occurred. As the interest in the underlying process is entirely on the change point τ, we may write the underlying process of interest as $\theta(t) = \left\{\mathbf{I}_{\{\tau \leq t\}}\right\}$. This paper is restricted to cases when a change occurs in the distribution of the variables from a known in-control distribution F^0 to a distribution F^1, the latter belonging to a finite family of known out-of-control distributions.

Many surveillance situations involve several, often dependent, variables, $\mathbf{X}(t) = (X_1(t), \ldots, X_p(t))$. If, in such multivariate situations, we expect a single change point τ where all the variables change simultaneously, the problem can be transformed into a univariate surveillance problem, see Wessman (1998). However, a multivariate surveillance situation typically involves surveillance of several processes where each component process may change at a different point. The multivariate surveillance problem thus generally involves several change points.

Consider for example the surveillance of product quality in a manufacturing process. The quality of a product is usually defined by numerous variables such as dimensions, breaking strength, numbers of flaws in the finish of the product etc. depending on the product. It is common to surveil several such variables. Depending on the manufacturing process conserned it may be suitable to assume that a change will occur simultaneously in all the variables or that it may occur only in one of several possible subsets of the surveilled variables.

Here the case where each variable is restricted to one change during the surveillance, is considered. Thus up to p change points, $(\tau_1, \ldots, \tau_p)$, are possible. For a specific surveillance situation some structural order of the occurrence of these changes might exist. Hawkins (1993) for example considers what he calls *a cascade type process* where a natural order in which the variables change exists. This case will not be discussed here.

For situations with several change points a definition may be necessary of when the multivariate process $\boldsymbol{\theta}(t) = (\theta_1(t), \ldots, \theta_p(t))$ is to be considered

out-of-control. In this paper the first change point occurring is considered as the change point of interest. That is, if any variable departs from its in-control distribution the process will be considered as out-of-control. Although we are only interested in detecting the first change, any subsequent change will influence the distribution of the observation process and thus the properties of a surveillance method. It is therefore of interest to include the following changes when comparing surveillance methods.

A number of multivariate surveillance methods based on well-known univariate surveillance methods such as the Shewhart, CUSUM and EWMA methods have been suggested so far. The adoption of these univariate methods to multivariate surveillance usually runs along two lines. According to the first approach, (a) parallel univariate surveillance methods are applied separately to each observed variable and an alarm is given when the first of these gives a signals. Multivariate surveillance methods based on parallel Shewhart have, been suggested by e.g. Timm (1996) and parallel CUSUM methods by among others Woodall and Ncube (1985). According to the second approach (b) the information from the different variables is first combined by a real-valued statistic. The statistic is then surveilled using some common univariate method. Examples of this type of surveillance method are the methods suggested by Hotelling (1947) and Crosier (1988) who suggested Shewhart and CUSUM methods respectively based on the T^2-statistic. This classification of methods is based on technical aspects of the form of the alarm function and does not take into account the inferential aspects on which amount of information used.

Another way to categorise multivariate surveillance methods is to consider which information they use at each decision point, how the method combines available information over time as well as how the method combines information from one single observation point. The former is of interest in surveillance in general but the latter is specific for multivariate surveillance. This paper considers the latter. For this reason the multivariate methods considered in this paper are all of the Shewhart type, that is, they base their decision on information only from the last observation point.

This simplifies the comparison between the two approaches (a) and (b) of combining information between variables. The methods considered here for combining information are the T^2-statistic discussed by, a.o. Alt (1985), and two types of finite intersection statistics, suggested by Hayter and Tsui (1994) and Hawkins (1991), and a surveillance method derived from the LR-procedure to construct surveillance methods described by Frisén and de Maré (1991).

The comparison is made for the case of simultaneous surveillance of two processes observed at discrete intervals through two correlated standard normal variables. The observations are independent over time and the covariance structure between variables is considered to be known. Each process is allowed to experience up to one change during the surveillance period.

In Section 2 the surveillance problem considered here is described. In Section 3 the surveillance methods considered are described. In Section 4 the results of the comparisons are presented and discussed. Finally, in Section 5, some concluding remarks are made.

2 Specification of the process

We consider a situation where the distribution of each observed variable, $X_i = \{X_i(t)\}_{t\in T}, i = 1,\dots,p$, can only change once, from a known in-control distribution, $F^0_{X_i}$, to a new known out-of-control distribution, $F^1_{X_i}$. The change points, $\tau_i, i = 1,\dots,p$, where these changes occur, are unknown random time points with some joint distribution G and marginal densities g_i defined on $T = \{1,2,\dots\}$. Given the change points, the observations are time independent observations, that is if $\mathbf{X}_i(T_l) = \{X_i(t) : t \in T_l\}$:

$$\mathbf{X}_1(T_l)\,|(\tau_1,\dots,\tau_p) \perp \mathbf{X}_2(T_r)\,|(\tau_1,\dots,\tau_p)\ ,\forall T_l, T_r \subset N, T_l \cap T_r = \varnothing, \quad i = 1,\dots,p.$$

Numerical comparisons are made for the bivariate case, $p = 2$, with variables that, given the change points (τ_1,τ_2), are distributed $(X_1(t), X_2(t)) \sim N_2(\boldsymbol{\mu}(t), \Sigma)$, $t = 1,2,\dots$ with covariance

$$\rho^2 = E\left[(X_1(t) - \mu_1(t))(X_2(t) - \mu_2(t))\right]$$

and unit variance $\sigma_{ii} = 1$. The comparison is restricted to cases with positive correlation, $\rho \geq 0$, the effect of negative correlation is only briefly discussed. Let $\tau_{(1)} = \min(\tau_1,\tau_2)$ and $\tau_{(2)} = \max(\tau_1,\tau_2)$. The family of joint densities possible for $(X_1(t), X_2(t))$ is in this case

$$\mathcal{F} = \begin{cases} F^{00} & t < \tau_{(1)} \\ F^{10} & \tau_1 \leq t < \tau_2 \\ F^{01} & \tau_2 \leq t < \tau_1 \\ F^{11} & \tau_{(2)} \leq t \end{cases} \tag{1}$$

where all included distributions are known and let f^{ij} be the corresponding densities. The unobserved processes of interest can thus be defined as $\boldsymbol{\theta}(t) = \left(\mathbf{1}_{\{\tau_1 \geq t\}}, \mathbf{1}_{\{\tau_2 \geq t\}}\right)$.

The case when a sudden shift in the mean from a known in-control level $\boldsymbol{\mu}^{00}$ to known out of control levels, $\boldsymbol{\mu}^{10}$, $\boldsymbol{\mu}^{01}$ or $\boldsymbol{\mu}^{11}$, is considered. Furthermore we write $\boldsymbol{\mu}^{ij} = \left(\mu_1^i, \mu_2^j\right)'$, $i, j = 0, 1$. When in the comparisons a specification of the distribution of the change points is necessary they are specified by a bivariate geometric distribution defined on $\{1, 2, \dots\} \times \{1, 2, \dots\}$ with intensities $(\nu_{10}, \nu_{01}, \nu_{11})$.

3 Methods

The primary goal in surveillance is to, on-line, detect the change points as fast as possible and with a minimum of false alarms. To our aid we have the observations made so far,

$$\mathcal{X}_t = (\mathbf{X}(1), \dots, \mathbf{X}(t)) = \left((X_1(1), \dots, X_p(1))', \dots, (X_1(t), \dots, X_p(t))'\right),$$

and any knowledge about the distribution of the observed variables. In our bivariate case the possible distributions for $\mathbf{X}(t)$ given the change points are $\mathcal{F} = \{F^{ij}; i, j = 0, 1\}$, and any àpriori information available on the distribution of the change points.

To detect a change a surveillance method is used to signal an alarm whenever sufficient evidence of a change exists. The surveillance method can be defined through a stopping rule

$$t_A = \min\{t \,|\, \mathcal{X}_t \in A_t \subseteq V_{\mathcal{X}_t}\}$$

where $V_{\mathcal{X}_t}$ is the set of all possible values of $\mathcal{X}_t$ and A_t, are the values of $\mathcal{X}_t$ for which alarms are made. Based on the observations made up to time s this stopping rule decides, after each new observation, if an alarm should be made or not.

The alarm region, A_t, is usually defined through an alarm function $p(\cdot)$ and a critical limit, $k(t)$, so that the surveillance method may be defined by

$$t_A = \min\{t : p(\mathcal{X}_t) > k(t)\}.$$

The critical limit determines the size of the alarm region and is usually set to give a certain false alarm rate. The false alarm rate is usually measured

by the average time to an alarm when no change occurs, ARL^0. This will be discussed further in Section 4.

The alarm function determines the shape of the alarm region and by that the method's properties in different situations. In our case the ability to detect shifts occurring in only one or in both of the variables is of interest.

The surveillance methods considered in this paper are all Shewhart type methods and can be defined by stopping rules of the form

$$\begin{aligned} t_A &= \min\left\{t : \mathbf{x}(t) \in A(t) \subseteq V_{\mathbf{x}(t)}\right\} \\ &= \min\left\{t : p(\mathbf{x}(t)) > k\right\}. \end{aligned}$$

Thus, they base their alarm functions only on the last observation made which give them alarm regions of a simple form. This makes them suitable for comparing different ways to combine the information from one observation time point.

3.1 Methods based on the marginal distributions

One approach to the construction of a multivariate surveillance method is to use the marginal distribution of the observed variables to construct the alarm function. The most common example of such a method is to apply separate surveillance methods for each variable and to signal an alarm when the first of these component methods signals.

A multivariate surveillance method constructed in this way is a special case of a type of method which is defined by the stopping rule

$$t_A = \min_{1 \le i \le r} [t_{A,i}]$$

where $t_{A,i} = \min\left\{t : p_i(\mathcal{X}_t) > k_i(t)\right\}, i = 1, \ldots, r$ are the $r > 1$ component surveillance methods. These component surveillance methods are specified to detect specific changes in the multivariate variable, usually changes in the individual variables. The methods considered have $r = p$ component methods, $t_{A,i}$, each constructed to detect a change in the i:th variable. However, the set of component methods could also include other combinations of the surveilled variables such as for example some or all of the principal components, as suggested by for example Jackson (1985). Reasons for applying this type of approach include of surveillance methods are usually simple to construct and that these methods have diagnostic properties regarding which of the component processes that has changed.

Recently multivariate surveillance methods based on the union intersection principle applied on the marginal distribution have been suggested by for example Hayter and Tsui (1994) and Timm (1996). They suggest methods based on separate parallel Shewhart methods where the component methods are based on the marginal distributions of the observed variables. Also, other multivariate surveillance methods of this type have been suggested. For example Woodall and Ncube (1985) suggested the use of parallel CUSUM methods.

Hayter and Tsui (1994) suggested a multivariate method for detecting changes in the mean of a multivariate normally distributed random variable based on the component surveillance methods

$$t_{A,i} = \min\left\{t : \frac{|X_i(t) - \mu_i^0|}{\sigma_{ii}} > k\right\}$$

where μ_i^0 is the in-control level and σ_{ii}^2 is the variance for the X_i. An alarm is signalled as soon as

$$t_A = \min\left\{t \left| \max_{1\leq i\leq p}\left[\frac{|X_i(t) - \mu_i^0|}{\sigma_{ii}}\right] > k\right.\right\}$$

and the critical limit is set so that the false alarm rate is controlled exactly at a given level p_{00}. It is the exact control of p_{00} that distinguishes their method from earlier suggested methods of this type. This method will be referred to as the $M(two)$ in the discussion below.

Notice that the random variables are standardized in the alarm functions and that the critical limits are equal. Thus, equal weights are given to all variables in t_A. This choice of identical critical limits is appropriate when we assume that all variables have the same intensity, $\nu(t) = P(\tau_i = t \,|\tau_i \geq t), i = 1, \ldots, p$ of a change and changes of the same magnitude, $|\mu_i^1| = \delta, i = 1, \ldots, p$, are expected in all variables.

Thus the alarm function excludes any knowledge of the correlation structure between observations made at the same time point and can be based on the marginal log likelihood ratios,

$$\log\left\{lr^i(\mathbf{x}(t))\right\} = \log\left\{f^1(x_i(t))/f^0(x_i(t))\right\}, i = 1, 2. \tag{2}$$

A one-sided version of this method to detect an increase in the mean values is also considered. Based on the statistic

$$\max_{1\leq i\leq p}\left[\frac{X_i - \mu_i^0}{\sigma_{ii}}\right] = \max_{1\leq i\leq p}\left[lr^i(\mathbf{X}(t))\right] \tag{3}$$

it signals at

$$t_A = \min\left(t \left| \max_{1\leq i\leq p} \left[lr^i\left(\mathbf{X}\left(t\right)\right)\right] > k^{\dagger}\right.\right). \tag{4}$$

Note that also here the critical limit is identical for all component processes. This method will be referred to as $M(one)$.

Timm (1996) noted that this is approach is equivalent with a sequential use of the single step Finite Intersection Test (FIT) of Krishnaiah (1979) where the familywise error (FWE) is controlled at a level $\frac{1}{\gamma}$. Timm therefore called this method the FIT-method. The FWE is the probability that at least one of the component surveillance methods signals a false alarm at a given time t. As, according to Hochberg and Tamhane (1987), any union intersection method on a finite set controlling familywise error (FWE) can be called an FIT method, this is not the only FIT method possible. For a more thorough discussion of FWE see a.o. Hochberg and Tamhane (1987).

3.2 Methods based on the joint distribution

A possible drawback of the method based on the marginal distributions is that the component methods, $t_{A,i}, i = 1, \ldots, r$, does not use the information of a change in variable X_i available in the other variables. Several multivariate surveillance methods with alarm functions based on the joint distribution which combine information from all variables has also been considered.

3.2.1 A parallel surveillance method

Hawkins (1991) suggests a method technically similar to the M-method of Hayter and Tsui (1994) where the component methods are based on the scaled residuals from the regression of each variable on all others, that is, he surveilled the transformed variable

$$\mathbf{Z}\left(t\right) = L\mathbf{X}\left(t\right)$$

where $L = \sqrt{diag\left(\Sigma^{-1}\right)^{-1}}\Sigma^{-1}$, Σ as before the covariance matrix, instead of the original variables $\mathbf{X}\left(t\right)$. Thus, each component alarm function is based on the joint distribution of the observed variables. Hawkins (1991) noted that this is the same as to base the component surveillance methods on the likelihood ratios.

For $p = 2$ these are

$$lr^{10}(\mathbf{x}(t)) = \frac{f^{10}(\mathbf{x}(t))}{f^{00}(\mathbf{x}(t))} \text{ and } lr^{01}(\mathbf{x}(t)) = \frac{f^{01}(\mathbf{x}(t))}{f^{00}(\mathbf{x}(t))}. \tag{5}$$

If the observed variables are distributed $\mathbf{X}(t) \stackrel{d}{=} N_2(\boldsymbol{\mu}^{00}, \Sigma)$, when the processes are in control, then the transformation becomes

$$\mathbf{Z}(t) = \frac{1}{\sqrt{1-\rho^2}} \begin{bmatrix} X_1(t) - \rho X_2(t) \\ X_2(t) - \rho X_1(t) \end{bmatrix}.$$

Hawkins (1991) suggests surveillance methods based on either parallel Shewhart or CUSUM methods for each Z_i. We shall consider a one-sided parallel Shewhart method based on $\mathbf{Z}$. The surveillance method considered is based on the stopping rule $t_A^* = \min\left(t_{A,1}^*, t_{A,2}^*\right)$ where

$$\begin{aligned} t_{A,1}^* &= \min\left\{t : \frac{X_1(t) - \rho X_2(t)}{\sqrt{1-\rho^2}} > k_z\right\} \\ t_{A,2}^* &= \min\left\{t : \frac{X_2(t) - \rho X_1(t)}{\sqrt{1-\rho^2}} > k_z\right\}. \end{aligned}$$

thus the stopping rule can alternatively be written $t_A^* = \left\{t : \max_{1 \le i \le 2}\left[Z_i(t)\right] > k_z\right\}$ or $t_A^* = \{t : \max\left[lr^{10}(\mathbf{x}(t)), lr^{01}(\mathbf{x}(t))\right] > k_z\}$. This method will hereafter be referred to as Hawkins method or $H(one)$.

3.2.2 The T^2-Method

The first multivariate surveillance method based on the joint distribution was probably the T^2-method of Hotelling (1947), which, for a known covariance matrix is based on the alarm statistic

$$p_{T^2}(\mathbf{x}(t)) = \left(\mathbf{x}(t) - \boldsymbol{\mu}^{00}\right)' \Sigma^{-1} \left(\mathbf{x}(t) - \boldsymbol{\mu}^{00}\right),$$

see for example Alt (1985). A number of other methods, based on the T^2-statistic, have since then been suggested. For example Alwan (1986) and Crosier (1988) have suggested the use of a CUSUM method based on this function.

The T^2-method is based on the stopping rule

$$t_A = \min\left\{t \,|p_{T^2}(\mathbf{X}(t)) > \chi^2_{p,\alpha}\right\} \tag{6}$$

where $\chi^2_{p,\alpha}$ is the upper α percentile for the χ^2-distribution with p df. In contrast to the previous methods, no specification of type or size of the shift is necessary, the T^2-method being based on a pure significance test. As noted by Hawkins (1991) there is a relation between the $H(one)$-method and the T^2-statistic. In the case considered we have

$$p_{T^2}(\mathbf{X}(t)) = \sum_{i=1}^{p} \frac{(X_i(t) - \mu_i^0) Z_i(t)}{\sqrt{1-\rho^2}}$$

thus the alarm function of the T^2-method can be written as a function of the likelihoods $lr^{10}(\mathbf{x}(t))$ (5) and the marginal likelihood ratios $lr^i(\mathbf{x}(t))$ (2).

3.2.3 Likelihood ratio methods

All the methods discussed here so far can be expressed by combinations of partial likelihoods. A method based on the full likelihood is the general likelihood ratio method discussed by Frisén and de Maré (1991) for univariate surveillance. It is based on the alarm function

$$LR(\mathcal{X}_t | C(t)) = \frac{dP(\mathcal{X}_t | C(t))}{dP(\mathcal{X}_t | D(t))}. \tag{7}$$

where $C(t)$ is the critical event of interest concerning (τ_1, τ_2) and $D(t) = \{\tau_{(1)} > t\}$. The method signals an alarm when

$$t_A = \min\{t | LR(\mathcal{X}_t | C(t)) > k(t)\}$$

where the critical limit may depend on time. Frisén and de Maré (1991) showed that a surveillance method constructed this way is optimal in the following sence:

Theorem 1 *For specified false alarm probability $P(\mathcal{X}_t \in A_t | D(t))$ the general likelihood ratio method maximises the probability $P(\mathcal{X}_t \in A_t | C(t))$ of detecting the critical event $C(t)$ of interest.*

In the univariate case the LR-method is equivalent to the ordinary Shewhart method when the critical event is specified as $C(t) = \{\tau = t\}$. If in the multivariate case the critical event is specified as $C(t) = \{\tau_{(1)} = t\}$, the alarm statistic reduces to

$$LR(\mathcal{X}_t | \tau_{(1)} = t) = \frac{dP(\mathbf{X}(t) | \tau_{(1)} = t)}{dP(\mathbf{X}(t) | \tau_{(1)} > t)}$$

and depends also here on the last observation only. The LR-method defined to detect $\{\tau_{(1)} = t\}$ is thus a Shewhart type method as defined here.

As this alarm function can be written as

$$\begin{aligned}\frac{dP\left(\mathbf{X}(t)\left|\tau_{(1)}=t\right.\right)}{dP\left(\mathbf{X}(t)\left|\tau_{(1)}>t\right.\right)} &= P\left(\tau_1=t,\tau_2>t\left|\tau_{(1)}=t\right.\right)\frac{f^{10}(\mathbf{X}(t))}{f^{11}(\mathbf{X}(t))}+\\ &\quad P\left(\tau_1>t,\tau_2=t\left|\tau_{(1)}=t\right.\right)\frac{f^{01}(\mathbf{X}(t))}{f^{11}(\mathbf{X}(t))}+\\ &\quad P\left(\tau_1=\tau_2=t\left|\tau_{(1)}=t\right.\right)\frac{f^{11}(\mathbf{X}(t))}{f^{11}(\mathbf{X}(t))}\end{aligned}$$

the LR-method for detecting $\{\tau_{(1)} = t\}$ can be written as

$$t_A = \min\left\{t\left|\sum_{i,j=0:(i,j)\neq(0,0)}^{1} w_{ij}(t)\, lr^{ij}(\mathbf{x}(t)) > k\right.\right\} \tag{8}$$

with the weights

$$\left\{\begin{array}{l} w_{10} = P\left(\tau_1=t,\tau_2>t\left|\tau_{(1)}=t\right.\right)\\ w_{01} = P\left(\tau_1>t,\tau_2=t\left|\tau_{(1)}=t\right.\right)\\ w_{11} = P\left(\tau_1=\tau_2=t\left|\tau_{(1)}=t\right.\right)\end{array}\right. .$$

The alarm function is thus a weighted sum of the possible likelihood ratios, $lr^{ij}(\mathbf{x}(t)) = f^{ij}(\mathbf{x}(t))/f^{00}(\mathbf{x}(t)), (i,j) \neq (0,0)$, and dependent on the distribution of the change points τ_1 and τ_2.

In certain situations the critical limit for the alarm function is made dependent on time, see Frisén and Wessman (1999), but it is here held constant over time, $k(t) = k$. The method (8) is in the following referred to as the $LR_{\mathbf{w}}$-method.

Note that the weights, $w_{ij}(t)$, are proportional to the probability that a change occurring at time t is in that specific direction. Thus, $w_{10}(t) > w_{01}(t)$ for example, indicates that a change in the first process is more likely at time t. The size of the third weight, $w_{11}(t)$, can be seen as reflecting the correlation between the two change points at t for which the method is optimal.

The LR-method is, if not otherwise stated, specified with the weights,

$$w^* = (w_{10}^*, w_{01}^*, w_{11}^*)' = \left(\sqrt{2}-1, \sqrt{2}-1, \left(\sqrt{2}-1\right)^2\right). \tag{9}$$

These weights correspond to choosing the weights as $w_{11} = w_{10}^2$. Two other choices are also briefly considered, $\mathbf{w} = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$ and $\mathbf{w} = \left(\frac{1}{4}, \frac{1}{4}, \frac{1}{2}\right)$, corresponding to specifying the weights as $w_{11} = w_{10}$ and $w_{11} = 2w_{10}$.

The weights $\mathbf{w}^*$ also correspond to the case with uncorrelated change points and equal marginal intensities $v \approx 0.3$. More generally, for given equal marginal densities, v, these weights correspond with a correlation between change points of

$$Corr\left(\tau_1, \tau_2\right) = \sqrt{\frac{\left(\frac{w_{10}^*}{\sqrt{2}} v - v^2\right) v}{2v - \frac{w_{10}^*}{\sqrt{2}} v}}.$$

Thus, these weights also correspond to cases with other marginal intensities. For example, if the marginal intensities are lower than 0.3 but still equal, then the weights correspond to a situation with positive correlation between the change points.

4 Results and discussion

4.1 The run length distribution

As the Shewhart type method depend only on the distribution of the last observation the alarm probabilities at a given time point are

$$P\left(t_A = t \middle| t_A \geq t\right) = \begin{cases} p_{00} & t < \tau_{(1)} \\ p_{10} & \tau_1 \leq t < \tau_2 \\ p_{01} & \tau_2 \leq t < \tau_1 \\ p_{11} & \tau_{(2)} \leq t \end{cases}.$$

where

$$\begin{cases} p_{00} = P\left(t_A = t \middle| t_A \geq t, \tau_{(1)} > t\right) \\ p_{10} = P\left(t_A = t \middle| t_A \geq t, \tau_1 \leq t, \tau_2 > t\right) \\ p_{01} = P\left(t_A = t \middle| t_A \geq t, \tau_1 > t, \tau_2 \leq t\right) \\ p_{11} = P\left(t_A = t \middle| t_A \geq t, \tau_{(2)} \leq t\right). \end{cases}$$

Thus, the run length distribution for the Shewhart type method is:

$$P\left(t_A = t\right) = \begin{cases} \left(1 - p_{00}\right)^{t-1} p_{00} & t < \tau_{(1)} \\ \left(1 - p_{00}\right)^{\tau_1 - 1} \left(1 - p_{10}\right)^{t - \tau_1} p_{10} & \tau_1 \leq t < \tau_2 \\ \left(1 - p_{00}\right)^{\tau_2 - 1} \left(1 - p_{01}\right)^{t - \tau_2} p_{01} & \tau_2 \leq t < \tau_1 \\ \left(1 - p_{00}\right)^{\tau_1 - 1} \left(1 - p_{10}\right)^{\tau_2 - \tau_1} \left(1 - p_{11}\right)^{t - \tau_2} p_{11} & \tau_1 \leq \tau_2 \leq t \\ \left(1 - p_{00}\right)^{\tau_2 - 1} \left(1 - p_{01}\right)^{\tau_1 - \tau_2} \left(1 - p_{11}\right)^{t - \tau_1} p_{11} & \tau_2 \leq \tau_1 \leq t \end{cases}.$$

All comparisons between methods can therefore be based on the alarm probabilities, p_{ij}. Note that the run length distribution is also dependent on the change points (τ_1, τ_2) and consequently on the distribution of these random variables. This is a difference from the univariate Shewhart method where the change points distribution does not influence the run length distribution.

4.2 Critical limits for the alarm functions

As mentioned in Section 2 the comparisons are made for the surveillance of, given (τ_1, τ_2), normally random distributed variables $(X_1(t), X_2(t)) \sim N_2(\boldsymbol{\mu}(t), \Sigma)$, $t = 1, 2, \ldots$, to detect a sudden shift in the mean vector. In all numerical comparisons we consider the case

$$\boldsymbol{\mu}(t) = \begin{cases} (0,0)' & t < \tau_{(1)} \\ (1,0)' & \tau_1 \leq t < \tau_2 \\ (0,1)' & \tau_2 \leq t < \tau_1 \\ (1,1)' & t \geq \tau_{(2)} \end{cases} \tag{10}$$

where the covariance matrix is known and fixed. Comparisons are made for unit variance and correlation $\rho = 0.0, 0.2, 0.4, 0.6$ and 0.8. To make the methods comparable the critical limits are set so that the average run length until an alarm when no change occurs is

$$\mathrm{ARL}^0 = E\left[t_A \left| \tau_{(1)} = \infty \right.\right] \geq \gamma$$

with $\gamma = 100$.

For the Shewhart type method this is equivalent to specifying the critical limits so that the false alarm probability is

$$p_{00} = P\left(\mathbf{X}(t) \in A(t) \left| \boldsymbol{\mu}(t) = \boldsymbol{\mu}^{00}\right.\right) \leq \frac{1}{\gamma}.$$

In our case $p_{00} = 1/100 = 0.01$.

4.2.1 Parallel methods

One approach to determining the critical limits, k, is to use the Bonferroni inequality and to specify the critical limits so that the false alarm probabilities for the component processes, $p_{0,i}$, satisfy $p_{0,i} = \frac{1}{p\gamma}, i = 1, \ldots, p$. This approach is conservative for all values of ρ. Another way is to control the false alarm probabilities so that the Slepian inequality Hochberg and Tamhane

(1987), or Dunn-Sidak inequality is satisfied, that is to specify k so that the component false alarm probabilities satisfy

$$p_{0,i} = 1 - \left(1 - \frac{1}{\gamma}\right)^{1/p}, i = 1, \ldots, p.$$

Note that this means that in practice the critical limits are specified as if $\rho = 0$. For all $\rho > 0$ both approaches yield ARL^0 greater than γ and thus a loss of efficiency. The difference between these two approaches is small in the case considered. The only conservative methods we consider are the one and two-sided marginal surveillance methods based on the Bonferroni limits such that $p_{0,i} = \frac{1}{2\gamma} = 0.005$. These methods will be referred to as $B(one)$ and $B(two)$.

To avoid the loss of efficiency the critical limit must be specified so that $\mathrm{ARL}^0 = \gamma$ for $\rho > 0$. To achieve this the full knowledge of the distribution of the observed variables must be used. This is done by for example Hayter and Tsui (1994), $M(one)$ and $M(two)$, and Hawkins (1991), $H(one)$. As noted by Hayter and Tsui (1994), the difference between the conservative limits and those where the correlation is utilised can be quite small but becomes more increased for high ρ and γ. This is confirmed in this case by the results in Table 1 and Figure 1.

Critical limits controlling the false alarm probability exactly can be found for the parallel methods either by numerical integration or Monte Carlo methods, see for example Hayter and Tsui (1994). Tables are available for some values of ρ and p_{00}, see Timm (1996). Also, for the case $p = 2$, subroutines exist for calculating percentiles for bivariate normal cumulative distribution. One such subroutine, the PROBBNRML function in SAS, is used here both for the marginal methods, $M(one)$ and $M(two)$, and Hawkins method, $H(one)$, which is based on the joint distribution.

For the one-sided marginal parallel method we have that the false alarm probability can be written as

$$\begin{aligned} p_{00} &= P_{00}\left(\{X_1 > k\} \cup \{X_2 > k\}\right) \\ &= 1 - P_{00}\left(\{X_1 \leq k\} \cap \{X_2 \leq k\}\right). \end{aligned} \tag{11}$$

and for the two-sided

$$\begin{aligned} p_{00} &= P_{00}\left(\{|X_1| > k\} \cup \{|X_2| > k\}\right) \\ &= 1 - P_{00}\left(\{|X_1| \leq k\} \cap \{|X_2| \leq k\}\right). \end{aligned}$$

where

$$P_{00}\left(\{|X_1| \leq k\} \cap \{|X_2| \leq k\}\right) =$$
$$= P_{00}\left(X_1 \leq k, X_2 \leq k\right) + P_{00}\left(X_1 \leq -k, X_2 \leq -k\right) -$$
$$P_{00}\left(X_1 \leq -k, X_2 \leq k\right) - P_{00}\left(X \leq k, X_2 \leq -k\right).$$

Thus, the critical limits in both cases can be determined by calculation of the cumulative distribution function of the bivariate standard normal distribution.

The transformed variable, $\mathbf{Z}(t)$, used by the $H(one)$-method of Hawkins, also follows the multinormal distribution, $\mathbf{Z}(t) \overset{d}{=} N(\boldsymbol{\mu}^{00}, \mathbf{B})$ with $B = L\sqrt{diag\left(\Sigma^{-1}\right)^{-1}} = \{\{1, -\rho\}, \{-\rho, 1\}\}$, before any change, but with the correlation $\rho_Z = -\rho_X$. Thus, the same approach as above can be used.

Calculations gave the following critical limits for $p_{00} = 0.01$:

Table 1. *Critical limits for $M(one)$, $M(two)$ and $H(one)$ such that* $\mathrm{ARL}^0 = 100$.

	$\rho = 0.0$	$\rho = 0.2$	$\rho = 0.4$	$\rho = 0.6$	$\rho = 0.8$	$\rho \longrightarrow 1.0$
$M(two)$	2.8062	2.8049	2.7996	2.7860	2.7522	2,5758
$B(one)$	2.5750	2.5750	2.5750	2.5750	2.5750	2.5750
$M(one)$	2.5750	2.5722	2.5647	2.5476	2.5091	2,3263
$B(two)$	2.8062	2.8062	2.8062	2.8062	2.8062	2.8062
$H(one)$	2.5750	2.5757	2.5758	2.5758	2.5758	2.5758

Note that when $\rho = 0.0$, $\mu_i^0 = 0$ and $\sigma_{ii} = 1$ we have that

$$X_i(t) = \frac{X_i(t) - \rho X_j(t)}{\sqrt{1-\rho^2}} = \frac{X_i - \mu_i^0}{\sigma_{ii}}.$$

Thus, $H(one)$ is here identical to $M(one)$ for $\rho = 0.0$. Also note that for $\rho = 0.4$ the critical limit of $H(one)$ approximately equals, up to the fourth decimal, the $\frac{1}{\gamma}$-limit for the univariate two-sided Shewhart method. For $\rho < 0$ we have that the critical limits for $H(one)$ and $H(two)$ are those of $M(one)$ and $M(two)$ for $|\rho|$ and vice versa. The critical limits for $B(one)$ and $B(two)$ being the as for all values of $\rho \geq 0$.

4.2.2 Methods based on one combined statistic

The T^2-method has, when no change has occurred, a central chi-square distribution with p df. The critical limit can thus be determined by

$$P_{00}\left(p_{T^2}\left(\mathbf{X}\right) > k\right) = 1 - P_{00}\left(\mathbf{X}'\Sigma^{-1}\mathbf{X} \leq \chi^2_{2,p_{00}}\right) = p_{00}.$$

The critical limit for the T^2-method is, in contrast to the *parallel* methods, not a function of the correlation between the observed variables and regardless of ρ the critical limit for e.g. $p_{00} = 0.01$ is $\chi^2_{2,\alpha} \approx 9.2103$. However, it should be noted that the alarm region for (X_1, X_2) is dependent on ρ.

Since the alarm function of the LR-method is a weighted sum of lognormal variables, its distribution is not easily calculated. Therefore, estimates of k were made by Monte Carlo methods using SAS/IML. These simulations gave for weights $\mathbf{w}^* = \left(\sqrt{2}-1, \sqrt{2}-1, \left(\sqrt{2}-1\right)^2\right)$ the following critical limits, $\widehat{k}$, for the alarm function

Table 2. *Simulations of the critical limit for the $LR_{\mathbf{w}}$-method.*

	$\rho = 0.0$	$\rho = 0.2$	$\rho = 0.4$	$\rho = 0.6$	$\rho = 0.8$
$\widehat{k}_{0.99}$	5.35	5.01	4.98	5.54	7.99

The critical limits in Table 2 are based on 1,800,000 simulations for each ρ. Based on the asymptotic normality of the order statistic the confidence interval for each limit is less than $k_{0.99} \pm 0.03$. This uncertainty in the determination of k results in that all measurements considered being presented with two decimals accuracy for the LR-method.

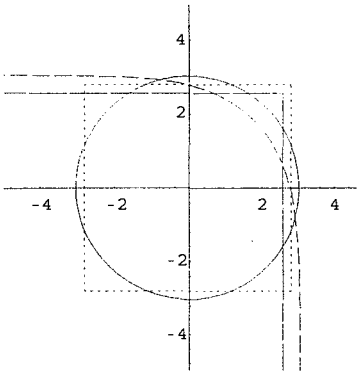
4.3 The alarm regions

Figure 1 shows the alarm regions based on the critical limits determined in Section 3.1.1 and 3.1.2 for $\rho = 0.0, 0.2, 0.4, 0.6$ and 0.8. The areas outside or above these contours being the alarm regions that give an ARL^0 of 100.

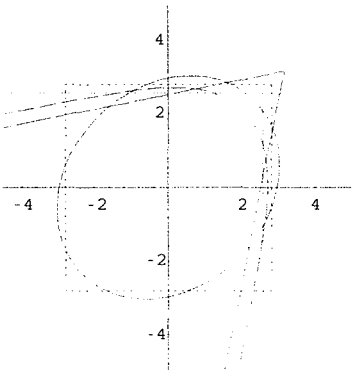
As can be seen there is a difference, both in shape and size, between the methods' alarm regions for a specific ρ. This indicates that the methods have different alarm probabilities for detecting a change in different directions. Of special interest in our comparison are differences between the methods' alarm regions in the direction of the main axes and along the line $X_1 = X_2$

as these directions represent the changes when only one process and both processes change respectively. The differences between the methods' alarm regions, showed in Figure 1 in these directions, correspond to differences in performances.

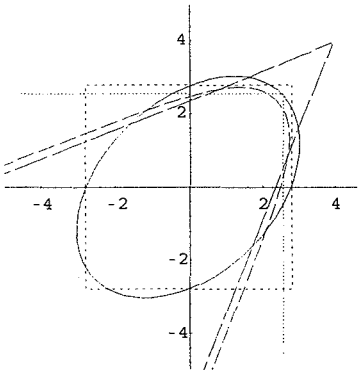
There is a difference between methods regarding the dependence on ρ. All except the marginal method have alarm regions highly influenced by the value of ρ. The marginal methods have similar alarm probabilities for detecting a certain type of shift, regardless of ρ. The $H(one)$-method based on the joint distribution coincides with the marginal $M(one)$-method for $\rho = 0$. For highly correlated data its alarm region is more close to the alarm region of the LR-method.



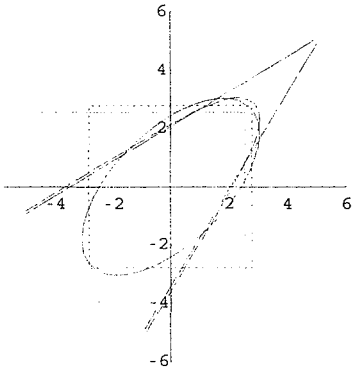
$\rho = 0$



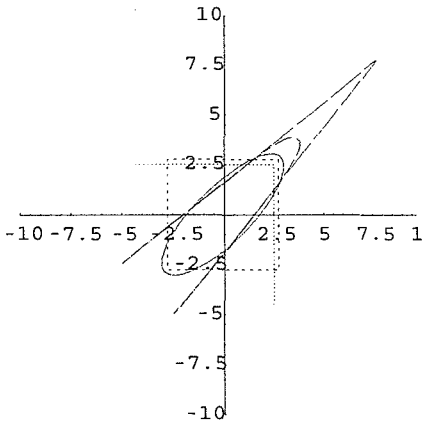
$\rho = 0.2$



$\rho = 0.4$



$\rho = 0.6$



$\rho = 0.8$

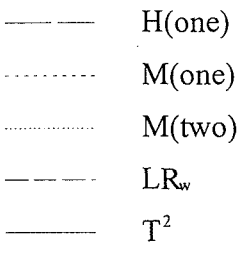


Figure 1. *Alarm regions in the* (X_1, X_2)*-plane,* $\mathrm{ARL}^0 = 100$.

4.4 The probability to detect a change

As described in Section 2 the aim is to detect a unit shift in the mean of one or both of the processes. Of interest is therefore the alarm probabilities p_{10}, p_{01} and p_{11} for detecting these shifts at a given time t. As both processes are surveilled for equal (unit) sized shifts and have the same variance we have that $p_{10} = p_{01}$. Note that, although these probabilities are constant over time for the Shewhart type method considered here, they are not so in general.

The alarm probabilities, p_{10} and p_{11}, for the one and two-sided marginal methods, $M(one)$ and $M(two)$, are found by numerical calculations as described above. For the Hawkins $H(one)$-method we can use the same relationship but with the difference that the shifts in the mean of the transformed variables, (Z_1, Z_2), is from $\boldsymbol{\mu}_z^{00}(t) = \mathbf{0}$ to

$$\boldsymbol{\mu}_z(t) = \begin{cases} \frac{1}{\sqrt{(1-\rho^2)}} \begin{pmatrix} 1 \\ -\rho \end{pmatrix} & \text{if } \boldsymbol{\mu}(t) = (1,0)' \\ \frac{1}{\sqrt{(1-\rho^2)}} \begin{pmatrix} -\rho \\ 1 \end{pmatrix} & \text{if } \boldsymbol{\mu}(t) = (0,1)' \\ \frac{1}{\sqrt{(1-\rho^2)}} \begin{pmatrix} 1-\rho \\ 1-\rho \end{pmatrix} & \text{if } \boldsymbol{\mu}(t) = (1,1)' \end{cases} .$$

The alarm function of the T^2-method follows a non-central chi-square distribution, $\chi^2\left(\delta^2\right)$, with 2 df and the non-centrality parameter

$$\delta^2\left(\boldsymbol{\mu}(t)\right) = \boldsymbol{\mu}(t)'\Sigma^{-1}\boldsymbol{\mu}(t) = \begin{cases} \frac{1}{1-\rho^2} & \text{if } \boldsymbol{\mu}(t) \in \{\boldsymbol{\mu}^{10}, \mu^{01}\} \\ \frac{2}{1+\rho} & \text{if } \boldsymbol{\mu}(t) = \boldsymbol{\mu}^{11} \end{cases}, \quad t = 1, 2, \ldots .$$

The alarm probabilities of the T^2-method are thus

$$P\left(p_{T^2}\left(\mathbf{x}(t)\right) > k^*\right) = \begin{cases} p_{10} = P\left(\chi^2\left(\frac{1}{1-\rho^2}\right) > \chi^2_{1-1/\gamma} \,|\boldsymbol{\mu}^{10}\right) \\ p_{11} = P\left(\chi^2\left(\frac{2}{1+\rho}\right) > \chi^2_{1-1/\gamma} \,|\boldsymbol{\mu}^{11}\right) \end{cases} .$$

For the $LR_{\mathbf{w}}$-method the alarm probabilities were determined by Monte Carlo methods. For each ρ, 10^6 simulations were made for each of p_{10} and p_{11}.

Table 3 shows the probability for the methods to immediately detect a change when the critical limits are set to give $\mathrm{ARL}^0 = 100$.

Table 3. *The alarm probabilities,* $\mathrm{ARL}^0 = 100$.

		$\rho = 0.0$	$\rho = 0.2$	$\rho = 0.4$	$\rho = 0.6$	$\rho = 0.8$
$B(one)$	p_{10}	0.0624	0.0619	0.0610	0.0598	0.0583
	p_{11}	0.1112	0.1086	0.1036	0.0965	0.0860
$H(one)$	p_{10}	0.0624	0.0626	0.0701	0.0929	0.1817
	p_{11}	0.1119	0.0780	0.0547	0.0379	0.0249
$M(one)$	p_{10}	0.0624	0.0622	0.0623	0.0631	0.0664
	p_{11}	0.1119	0.1092	0.1056	0.1016	0.0971
$M(two)$	p_{10}	0.0404	0.0403	0.0404	0.0410	0.0433
	p_{11}	0.0698	0.0683	0.0663	0.0638	0.0609
$B(two)$	p_{10}	0.0403	0.0402	0.0398	0.0392	0.0383
	p_{11}	0.0698	0.0681	0.0654	0.0612	0.0545
T^2	p_{10}	0.0404	0.0420	0.0478	0.0636	0.1268
	p_{11}	0.0845	0.0684	0.0577	0.0502	0.0446
$LR_{\mathbf{w}}$	p_{10}	0.06	0.06	0.07	0.09	0.18
	p_{11}	0.17	0.13	0.10	0.06	0.03

Table 3 shows that if a change has occurred in both processes all methods have a lower probability to detect it if the observations are highly correlated. If only one process has experienced a change, the situation is the reverse; then the probability of immediately detecting a change is higher in situations with high correlations between variables. An explanation to this is that the available information of a shift occurring in both variables decreases for highly correlated data, while the reverse is true if a shift occurs in only one variable. In fact for $\rho > 0.5$, the Mahalanobis squared distance, $d(\mathbf{x}) = (\mathbf{x} - \boldsymbol{\mu}^{00})' \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}^{00})$, of a unit shift in both processes is here smaller than for a unit shift occurring in only one process. This affects the likelihood ratios, $lr^{ij}(\mathbf{x}(t)), i, j = 0, 1$, but not the marginal likelihood ratios, $lr^{i}(x_i(t)), i = 0, 1$. For the marginal methods, the differences in the alarm probabilities are thus small but for the T^2 and the LR-method this effect is major. For the Hawkins $H(one)$-method the effect of the correlation between the variables is even more pronounced.

For $\rho < 0$ the alarm probabilities p_{10} and p_{11} increase as ρ decreases for all methods. This is specially so for the methods based on the joint distribution. The large decrease in p_{11} for these methods when $\rho \gg 0$ are substituted by a large increase when $\rho < 0$ decreases. The H, T^2 and $LR_{\mathbf{w}}$-methods can therefore be expected to perform better than the marginal M-methods for negatively correlated variables.

Although the $LR_{\mathbf{w}}$-method, according to Theorem 1, maximizes the expected alarm probability, $P\left(\mathbf{X}\left(t\right)\left|\tau_{(1)}=t\right.\right)$, for fixed false alarm probability, this is not so for p_{10} and p_{11} as can be seen in Table 3.

However, the $LR_{\mathbf{w}}$-method is also based on specifications of the distribution of the change points. These are reflected in the alarm functions weights which influence how the power is partitioned between the specified direction of interest. In Table 4a and 4b we consider $LR_{\mathbf{w}}$-methods giving more weight to the occurrence of a simultaneous change in both processes or larger intensity v (see Section 3.2.3).

Table 4a. *Alarm probabilities for the LR-method,* $\rho = 0.8$.

(w_{10}, w_{01}, w_{11})	p_{10}	p_{11}
$\mathbf{w}^*$	0.18	0.03
$\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$	0.17	0.04
$\left(\frac{1}{4}, \frac{1}{4}, \frac{1}{2}\right)$	0.16	0.06

Table 4b. *Alarm probabilities for the LR-method,* $\rho = 0$.

(w_{10}, w_{01}, w_{11})	p_{10}	p_{11}
$\mathbf{w}^*$	0.06	0.17
$\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$	0.06	0.18
$\left(\frac{1}{4}, \frac{1}{4}, \frac{1}{2}\right)$	0.06	0.18

In Table 4a we see that specifying a larger w_{11}-weight results in better probability to detect a change occurring in both processes, p_{11}. In fact for $w_{11} = 0.5$ this alarm probability is approximately that of the one-sided marginal parallel method for $\rho = 0.8$. For the case with uncorrelated variables in Table 4b the effect of using different weights is much smaller, but the same pattern appears.

Another way of measuring the alarm probability is to consider the probability to detect a specified critical event

$$p_w\left(t\right) = P\left(\mathcal{X}_t \in A_t \left| C\left(t\right)\right.\right) = E_{\tau_{(2)},,\tau_{(p)}}\left[P\left(t_A = t\right)\left|\tau_{(1)} = t, \tau_{(2)}, \ldots, \tau_{(p)}\right.\right].$$

According to Theorem 1, the LR-method is, regardless of the value of ρ, the method that maximises this alarm probability, if the weights are correctly

specified. In our case the critical event is chosen as $C(t) = \{\tau_{(1)} = t\}$ which gives

$$p_w(t) = P\left(t_A = t \,|\tau_{(1)} = t\right) = \sum_{i,j\in\{0,1\}} w_{ij}p_{ij}.$$

Thus, $p_{\mathbf{w}}$ is an expected alarm probability that gives weight to the alarm probabilities, p_{10}, p_{01} and p_{11}, according to how likely a specific change is. Table 5 gives these weighted alarm probabilities when the weights are $\mathbf{w}^* = \left(\sqrt{2}-1, \sqrt{2}-1, (\sqrt{2}-1)^2\right)$.

Table 5. *Expected alarm probabilities when* $\mathbf{w} = \mathbf{w}^*$.

	$\rho = 0.0$	$\rho = 0.2$	$\rho = 0.4$	$\rho = 0.6$	$\rho = 0.8$
$H(one)$	0.0709	0.0653	0.0675	0.0835	0.1548
$B(one)$	0.0706	0.0699	0.0683	0.0661	0.0630
$M(one)$	0.0709	0.0703	0.0697	0.0697	0.0717
$B(two)$	0.0454	0.0450	0.0442	0.0430	0.0411
$M(two)$	0.0454	0.0451	0.0448	0.0449	0.0463
T^2	0.0480	0.0465	0.0495	0.0613	0.1127
$LR_{\mathbf{w}^*}$	0.08	0.07	0.07	0.09	0.16
$LR_{(\frac{1}{3},\frac{1}{3},\frac{1}{3})}$	0.08				0.15
$LR_{(\frac{1}{4},\frac{1}{4},\frac{1}{2})}$	0.08				0.14

The $H(one)$, T^2 and $LR_{\mathbf{w}^*}$ methods, based on the joint distribution, do not suffer the loss of performance for high positive correlations in $p_{\mathbf{w}}$ as they do on p_{10} and p_{11}. The $LR_{\mathbf{w}^*}$-method has, as expected, the highest expected alarm probabilities, $p_{\mathbf{w}^*}$. But note that, even though the $LR_{(\frac{1}{3},\frac{1}{3},\frac{1}{3})}$ and $LR_{(\frac{1}{4},\frac{1}{4},\frac{1}{2})}$ have lower values, they still compare well with the other methods. The dual nature of the $H(one)$ method is once more visible in that it has the same value as the $M(one)$ for $\rho = 0$, but as ρ increases it converges to the $LR_{\mathbf{w}^*}$-method. Also, the difference between "one-sided" and undirected methods are more pronounced.

The alarm probabilities discussed so far reflect the method's power to detect a change immediately; another interesting property is the ability to detect a change within a certain time. This is measured by the Probability of a Successful Detection, PSD, suggested by Frisén (1992). The PSD measures

the ability to detect a change within a certain time and can in our case be defined as

$$\begin{aligned} \mathrm{PSD}\left(\tau_{(1)}, d\right) &= E_{\tau_{(2)}}\left[t_A \in [t, t+d) \,\middle|\, t_A \geq t, \tau_{(1)} = t, \tau_{(2)}\right] \\ &= \sum_{i=t}^{t+d-1} E_{\tau_{(2)}}\left[t_A = i \,\middle|\, t_A \geq t, \tau_{(1)} = t\right]. \end{aligned}$$

Thus, in contrast to the univariate case, the $\mathrm{PSD}\left(\tau_{(1)}, d\right)$ is dependent on the distributions of the second change point. For Shewhart type methods the PSD is dependent on d but not on t and we can write

$$\begin{aligned} \mathrm{PSD}\,(d) &= E_{\tau_{(2)}}\left[t_A \in [1, 1+d) \,\middle|\, t_A \geq 1, \tau_{(1)} = 1, \tau_{(2)}\right] \\ &= \sum_{i=1}^{d} E_{\tau_{(2)}}\left[t_A = i \,\middle|\, t_A \geq 1, \tau_{(1)} = 1, \tau_{(2)}\right] \end{aligned}$$

thus PSD(d) is still dependent on $\tau_{(2)}$. Alternatively, we can write PSD(d) as

$$\begin{aligned} \mathrm{PSD}\,(d) = \sum_{i=1}^{d-1} & P\left(\tau_{(2)} = i \,\middle|\, t_A \geq 1, \tau_{(1)} = 1\right) PSD\,(d\,|i) + \\ & P\left(\tau_{(2)} \geq d \,\middle|\, t_A \geq 1, \tau_{(1)} = 1\right) PSD\,(d\,|d) \end{aligned}$$

where $\mathrm{PSD}(d\,|i) = P\left(t_A \in [1, d) \,\middle|\, t_A \geq 1, \tau_{(1)} = 1, \tau_{(2)} - \tau_{(1)} = i\right)$.

Figure 2 shows $\mathrm{PSD}\left(d \,\middle|\, \tau_{(2)} - \tau_{(1)}\right)$ for $\tau_{(2)} - \tau_{(1)} = 1, 2, \ldots, 10$. The graphs illustrate that for the marginal methods there is a lower probability to detect $\tau_{(1)}$ when the second change occurs late, regardless of ρ. This is also the case for the $LR_{\mathbf{w}^*}$ and T^2-methods for variables with low correlation. For variables with high correlation there is instead an increase in probability for these methods. Notice that expected alarm probability p_{w^*} in Table 5 corresponds to PSD(d).

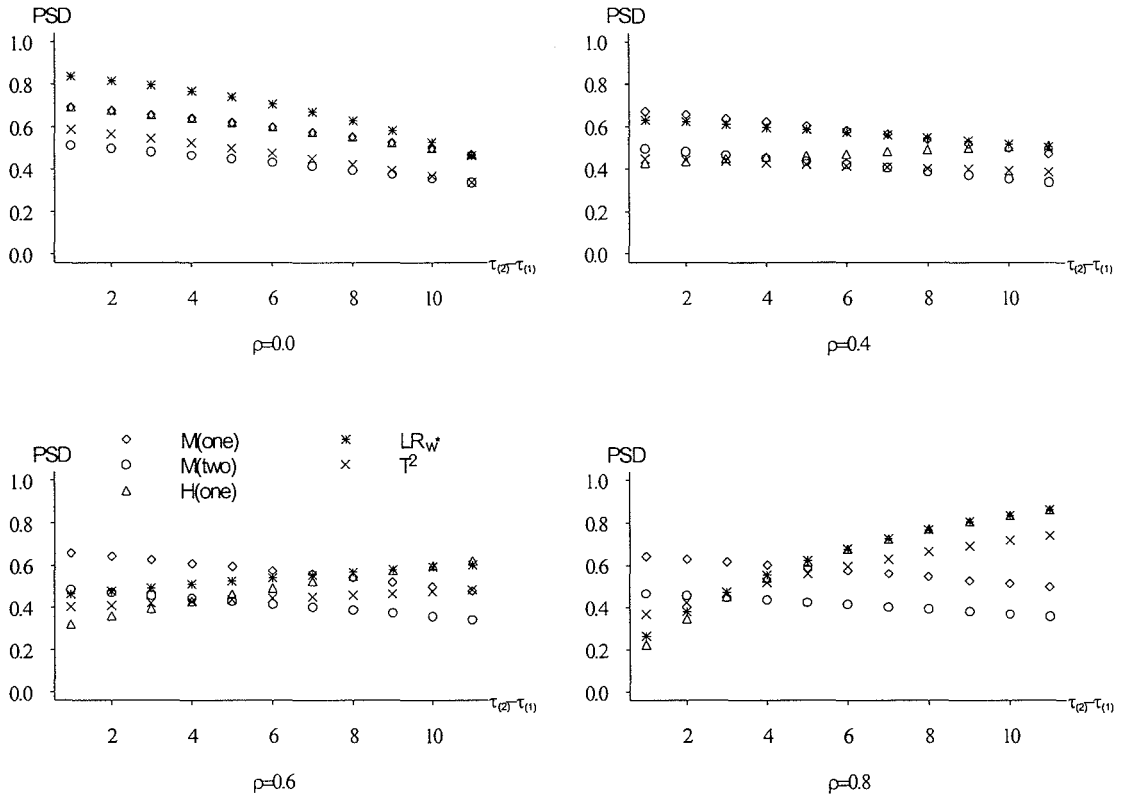


Figure 2. *The probability of an successful detection within 10 units.*

4.5 Average run length

The most commonly used measure of performance is the expected time to an alarm. The expected time to an alarm when a change occurs at the start is usually called ARL^1, in correspondence with ARL^0. For a univariate Shewhart method the ARL^1 is simply $1/p_1$, where p_1 is the probability to detect a change. When several change points are possible, the average run length to an alarm can be defined as

$$\mathrm{ARL}^1 = E\left[t_A \left| \tau_{(1)} = 1\right.\right].$$

and, as for PSD, ARL^1 is now dependent on the distribution of the change points.

A measure independent of (τ_1, τ_2) is ARL^1 given $\tau_{(2)}$ or $\mathrm{ARL}(\tau_1, \tau_2) =$

$E[t_A | \tau_1, \tau_2]$. Two special cases are

$$\begin{cases} \text{ARL}(1,1) = E\left[t_A \left| \tau_{(1)} = \tau_{(2)} = 1\right.\right] \\ \text{ARL}(1,\infty) = E\left[t_A \left| \tau_{(1)} = 1, \tau_{(2)} = \infty\right.\right] \end{cases}$$

that is the average run length when both processes change at the start, and when only one experiences a change at the start, the other not changing at all. These are the measures used by Hawkins (1991) in his comparison. In the case considered here these are:

Table 6. ARL$(1,1)$ *and* ARL$(1,\infty)$.

	ARL	$\rho = 0.0$	$\rho = 0.2$	$\rho = 0.4$	$\rho = 0.6$	$\rho = 0.8$
$H(one)$	$(1,\infty)$	16.04	15.96	14.26	10.77	5.50
	$(1,1)$	8.93	12.81	18.29	26.38	40.11
$M(one)$	$(1,\infty)$	16.05	16.07	16.05	15.84	15.07
	$(1,1)$	8.93	9.16	9.47	9.84	10.29
$B(one)$	$(1,\infty)$	16.04	16.16	16.42	16.39	17.17
	$(1,1)$	8.93	9.21	9.65	10.36	11.63
$M(two)$	$(1,\infty)$	24.79	24.79	24.72	24.37	23.11
	$(1,1)$	14.33	14.64	14.64	15.08	16.41
$B(two)$	$(1,\infty)$	24.79	24.86	25.10	25.52	26.09
	$(1,1)$	14.33	14.68	15.30	16.35	18.35
T^2	$(1,\infty)$	24.76	23.83	20.93	15.72	7.89
	$(1,1)$	11.84	14.62	17.33	19.92	22.40
LR	$(1,\infty)$	16.31	16.48	14.74	11.01	5.51
	$(1,1)$	6.01	7.79	10.51	16.38	32.36

For the Shewhart type method ARL$(1,1)$ and ARL$(1,\infty)$ are geometrically distributed with intensities p_{10} and p_{11}. The information found in Table 6 are the same as in Table 3. Though the large differences in performance for methods based on the joint distribution, for different ρ, become even more visible. The use of exact critical limits in $M(one)$ and $M(two)$ compared to the conservative Bonferroni limits of $B(one)$ and $B(two)$ gives marginally shorter ARL for weakly correlated data. For high positive correlations the gain is somewhat larger.

A limitation by restricting the comparison to just ARL$(1,1)$ and ARL$(1,\infty)$ is that we do not consider the behavior of the methods when a second change point occurs. To measure this we can, as for the PSD, consider $ARL^1\left(1, \tau_{(2)}\right), \tau_{(2)} = 1, 2, \ldots$, which starts at ARL$(1,1)$ and converges to ARL$(1,\infty)$ as $\tau_{(2)}$ increases.

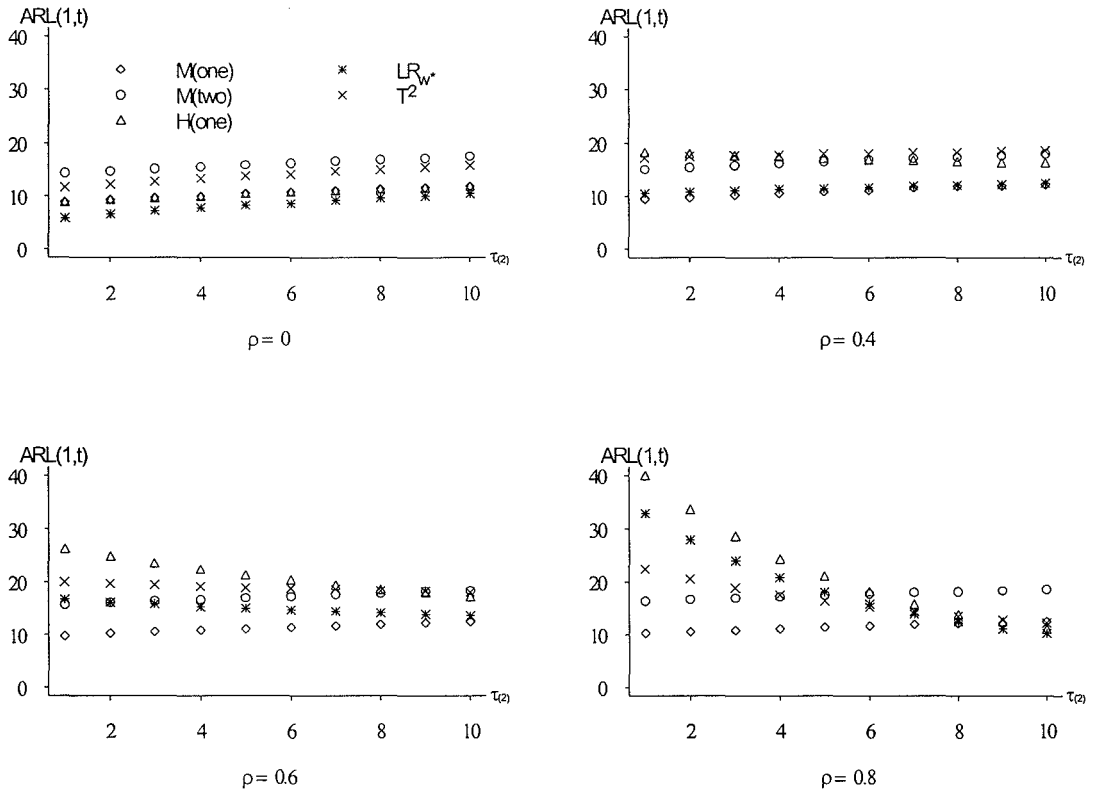


Figure 3. *Average Run length to motivated alarm when* $\tau_{(1)} = 1$.

Figure 3 shows $\mathrm{ARL}\left(1, \tau_{(2)}\right)$, $\tau_{(2)} = 1, , 10$. For independent variables or variables with low correlation the ARL increases with $\tau_{(2)}$ for all methods. The same tendency as in Figure 2 emerges, for highly positive correlations with a very high average run length specially for the $H(one)$ and $\mathrm{LR}_{\mathbf{w}^*}$-methods when the change points occur close in time. The shift in power from $\boldsymbol{\mu}^{11}$ to $\boldsymbol{\mu}^{10}$ for the methods based on the joint distribution of $\mathbf{X}(t)$ as ρ increases is also visible. Thus, which of the methods that has the best ARL depends both on the correlation between variables and the time before the second change can be expected to occur.

As seen above the ARL^1 depends on the distribution of the change points and can, if $G(t_1, t_2) = G(t_1, t_2)$ and $\mathrm{ARL}(t_1, t_2) = \mathrm{ARL}(t_2, t_1)\ \forall (t_1, t_2) \in \mathbb{N}^2$, be written

$$\mathrm{ARL}^1 = E\left[t_A \left| \tau_{(1)} = 1\right.\right] = \sum_{i=1}^{\infty} P\left(\tau_{(2)} = i \left| \tau_{(1)} = 1\right.\right) \mathrm{ARL}(1, i).$$

Another measure of performance is the expected delay until an alarm. Of interest are the conditional expected delay, $E[t_A - \tau^* | \tau^* = t, \tau^* \geq t_A]$, and the total expected delay, $E[t_A - \tau^* | \tau^* \geq t_A]$, where τ^* is the change point in a univariate surveillance situation. For the multivariate surveillance situation considered above where the first of several change points is of interest, these definitions generalise to $E\left[t_A - \tau_{(1)} \middle| \tau_{(1)} = t, \tau_{(1)} \geq t_A\right]$ and $E\left[t_A - \tau_{(1)} \middle| \tau_{(1)} \geq t_A\right]$. For the Shewhart type method we have that the conditional expected delay is independent of $\tau_{(1)}$. Thus, we have that

$$\begin{aligned} E\left[t_A - \tau_{(1)} \middle| \tau_{(1)} = t, \tau_{(1)} \geq t_A\right] &= \\ E\left[t_A - 1 \middle| \tau_{(1)} = 1, \tau_{(1)} \geq t_A\right] &= \mathrm{ARL}^1 - 1 \end{aligned}$$

and

$$E\left[t_A - \tau_{(1)} \middle| \tau_{(1)} \geq t_A\right] = E\left[t_A - 1 \middle| \tau_{(1)} \geq t_A\right] = \mathrm{ARL}^1 - 1.$$

Thus, the expected delay for a Shewhart type method is, as in the univariate case, a function of ARL^1.

5 Concluding remarks

The major difference in the considered method's performance seems to be whether their alarm function is based on the joint distribution or the marginal distributions of $\mathbf{X}(t)$. The performance of methods using alarm functions based on the joint distribution are highly influenced by the size of the correlation between the observed variables, ρ, whereas a method based on the marginal distribution has similar properties, regardless of the correlation between variables and the occurrence of a second change point.

Furthermore, for the methods based on the joint distribution the occurrence of the second change point, $\tau_{(2)}$, has for high correlations a major effect on both their alarm probabilities, p_{10}, p_{11}. In these situations, a simultaneous change or changes close together in time result in a bad performance for these methods in for example $\mathrm{ARL}(1, d)$, compared to the marginal methods. If the second change point occurs with some delay, large d, the performance of the joint distribution methods improves. This loss of alarm probability depending on when the second change occurs can be quite substantial. Thus substantial losses in performance are possible if available information of the distribution of the change points is not used.

This effect of high correlation is not reflected in the expected alarm probability. A method based on the joint distribution of $\mathbf{X}(t)$ performs better than the marginal methods, regardless of ρ. The method based on Hawkin's regression adjusted variables performs well for high ρ where it has nearly optimal expected alarm probabilities.

As can be seen in the comparison between the marginal methods, $M(one)$ and $M(two)$, substantial gain in performance can be made by using one-sided surveillance methods in situations were change in only one direction for each variable is of interest.

The choice of an appropriate method for combining information for a certain situation depends on the measure of optimality and the properties of the process.

References

[1] Alwan, L. C. (1986). CUSUM quality control - multivariate approach. *Communications in Statistics - Theory and Methods* **15,** 3531-3543.

[2] Alt, F. B. (1985). Multivariate Quality Control. In S. Kotz and N. L. Johnson Eds., *Encyclopedia of the Statistical Sciences*, vol. 6, 110-122. John Wiley & Sons, New York.

[3] Crosier, Ronald B. (1988). Multivariate Generalizations of Cumulative Sum Quality-Control Schemes. *Technometrics* **30**, 291-303.

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

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

[6] Frisén, M. and Wessman, P. (1999). Evaluations of likelihood ratio methods for surveillance, differences and robustness. *To appear in Communications in Statistics, Theory and Methods*, **28**.

[7] Hawkins, D. M. (1991). Multivariate Quality Control Based on Regression-Adjusted Variables. *Technometrics* **33**, 61-75.

[8] Hawkins, D. M. (1993). Regression Adjustment for Variables in Multivariate Quality Control *Journal of Quality Technology* **25**, 170-182.

[9] Hayter, A. J. and Tsui, K. L. (1994). Identification and Quantification in Multivariate Quality Control Problems. *Journal of Quality Technology* **26**, 197-208.

[10] Healy J. D. (1987). A Note on Multivariate CUSUM Procedures. *Technometrics* **29**, 409-412.

[11] Hochberg, Y. and Tamhane, A. C. (1987). *Multiple Comparison Procedures.* John Wiley & Sons, New York.

[12] Hotelling, H. (1947). Multivariate Quality Control. In C. Eisenhart, M. W. Hastay and W. A. Wallis Eds., *Technique of Statistical Analysis.* 111-184. McGraw-Hill, New York.

[13] Jackson, J. E. (1985) Multivariate quality control. *Communications in Statistics - Theory and Methods* **14**. 2657-2688.

[14] Krishnaiah, P. R. (1979). Some Developments on simultaneous Test Procedures. In P. R. Krishnaiah Eds., *Developments of Statistics*, vol 2. New York: North Holland.

[15] Timm, N. H. (1996). Multivariate Quality Control Using Finite Intersection Tests. *Journal of Quality Technology* **28**, 233-243.

[16] Wessman, P. (1998). Some principles for surveillance adopted for multivariate processes with a common change point. *Communications in Statistics, Theory and Methods* **27**, 1143-1161.

[17] Woodall, W. H. and Ncube, M. M. (1985). Multivariate CUSUM Quality-Control Procedures. *Technometrics* **27**, 285-292.

Research Report

1998:6	Dahlbom, U.:	Least squares estimates of regression functions with certain monotonicity and concavity/convexity restrictions.
1998:7	Dahlbom, U.:	Variance estimates based on knowledge of monotonicity and concavity properties.
1998:8	Grabarnik, P. & Särkkä, A.	Some interaction models for clustered point patterns.
1998:9	Afsarinejad, K. & Hedayat, S.:	Repeated measurement designs for models with self and mixed carryover effects.
1998:10	Hatemi-J, A. & Shukur, G.:	The causal nexus of government spending and revenue in Finland: A bootstrap approach.
1998:11	Shukur, G.:	The robustness of the systemwise Breauch--Godfrey autocorrelation test for non-normal distributed error terms.
1999:1	Andersson, E.:	On monotonicity and early warnings with applications in economics.