Research Report
Department of Statistics
Göteborg University
Sweden

Turning point detection using non-parametric statistical surveillance

Evaluation of some influential factors

Eva Andersson

Research Report 2001:3
ISSN 0349-8034

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

TURNING POINT DETECTION USING NON-PARAMETRIC STATISTICAL SURVEILLANCE

Evaluation of some influential factors

E. Andersson

Department of Statistics
Göteborg University
SE 405 30 Göteborg, Sweden

ABSTRACT

Turning point detection is important in many areas. One application is forecasting the time of the next turn in the business cycle, by detection of a turn in leading economic indicators. Another application is detection of a peak in the human menstrual cycle. In both these applications we make continual observation of the time series with the goal of detecting the turning point in the cycle as soon as possible. At each time, an alarm statistic and alarm limits are used in making a decision as to whether the time series has reached a turning point. The alarm statistic and the alarm-limit are based on the maximum likelihood ratio technique for surveillance. No parametric function is assumed for the cycle, but a non-parametric estimation procedure is used.

The shape of the turning point has an impact on the performance of the method for turning point detection. The influence of some turning point characteristics (slopes and smoothness of curve) is evaluated both theoretically and by simulation studies. The simulations are used to demonstrate the effect of the slopes (both pre-peak and post-peak). Results from the simulation study show that the false alarm probability increases for a non-smooth curve and that the detection probability is sensitive to the shape of the curve just around the turning point. In the theoretical investigation, it is shown that the expected delay of an alarm is shorter for a steeper post-peak slope. The method is also evaluated by applying it to a set of Swedish data.

Key words:Turning point detection; Non-parametric regression; Monitoring; Monotonic regression; Unimodal regression; Statistical surveillance; Business cycle.

1 INTRODUCTION

In analysis of cyclical processes, where the cycles are not periodic, it is often of interest to detect the turning points. Examples of areas where turning point detection is important are business cycle prediction, see Neftci (1982), Zarnowitz and Moore (1982), Jun and Joo (1993), Birchenhall et al. (1999, Hamilton (1989) and peak detection in biological cycles, see Royston (1991). In this report the methodology of statistical surveillance is used for turning point detection. Even though the technique is applicable for several applications, we will use the detection of a turn in a leading business indicator as an example in the following.

With statistical surveillance we mean continual observation of a process with the goal of detecting an important change in the underlying process as soon as possible. For a general review of statistical surveillance, see Shiryaev (1963), Frisén and de Maré (1991), Wetherill and Brown (1991), Srivastava and Wu (1993), Frisén and Wessman (1999). Since time is an important issue in surveillance, measures of performance which take into account the timeliness are used, rather than ordinary forecasting measures. Examples of such measures are the average run length (ARL), the expected delay (ED), the probability of successful detection (PSD), see Frisén (1992).

Generally, the problem of detecting a change in the process under surveillance can be presented as discriminating between two events, namely C (the change has occurred) and D (the change has not occurred). The discrimination is made using an alarm statistic together with alarm limits. The construction of the alarm system can be made in different ways, depending on the desired properties of the system. One method of surveillance is the likelihood ratio method (hereafter the LR method). For a fixed false alarm rate and a fixed time, the LR method has the highest probability of calling an alarm when the process has changed from one state to another. The LR method is optimal for detecting all changes before s (the time of decision) in the sense that the expected utility, based on the gain of an alarm and the loss of a false alarm, is maximized, see Frisén and de Maré (1991). The properties of the LR method in the situation where the process changes from an in-control level to an out-of-control level have been evaluated by Frisén and Wessman (1999).

For situations where the parameters of C and D are known, the LR method is optimal, but in the turning point detection situation, where the classes of distributions under C and D are composite, the LR method cannot be used. Frisén (1994) suggested the use of the maximum likelihood ratio as an alarm statistic for detecting turning points. This approach includes a non-parametric estimation procedure, see Frisén (1986), using only monotonicity restrictions. A surveillance situation where the aim is to detect a monotonic level drift from a constant level, using monotonicity restrictions, is presented in Arteaga and Ledolter (1997). The alarm statistic, based on the maximum likelihood ratio and suggested by Frisén (1994), was evaluated by Andersson (1999) for the case of a symmetrical peak with constant absolute growth. A turning point has many characteristics: symmetry, smoothness, sharp or non-sharp peak. In business cycles the peaks are often characterized by long expansion phases and shorter, steeper recession phases. The shape of the turning point has an impact on the ability to detect it. Simulations are made for four different peak shapes, where one peak is data on the Swedish industrial production and the other three peaks are modeled to mimic that peak. The three peaks were modeled using linear functions (with a sharp peak) and trigonometric functions (with a non-sharp peak). The median delay time, the expected delay time of an alarm and the probability of successful

detection are used as measures of the performance of the method based on the maximum likelihood ratio for the different peaks.

The paper is organized as follows: In Section 2 the model and surveillance system is presented. Section 3.1 contains a theoretical investigation of the influence of the post-peak slope. In Section 3.2 we present the results of a simulation study, concerning the influence of the shape of the turning point. Section 4 contains the result of an out-of-sample performance on Swedish data and a discussion on different measures of performance. Finally, Section 5 contains a discussion.

2 THE MAXIMUM LIKELIHOOD RATIO BASED METHOD

2.1 Model

We start the model discussion by studying some of the models used in previous research on turning point detection. The variable under surveillance is denoted by X.

Neftci (1982) assumes the following model for the increments of an observed economic time series

$$\Delta X(t) = \begin{cases} u^1 + \varepsilon^1(t), & t < \tau \\ u^2 + \varepsilon^2(t), & t \geq \tau \end{cases}$$

where u^1 and u^2 are constants,

$\varepsilon^1 \sim$ iid, $E[\varepsilon^1]=0$, $Var[\varepsilon^1] = \sigma_1^2$,

$\varepsilon^2 \sim$ iid, $E[\varepsilon^2]=0$, $Var[\varepsilon^2] = \sigma_2^2$, ε^1 is independent of ε^2.

The densities for the two regimes are estimated from previous data.

The model used by Layton (1996) for the possibly differentiated series is

$$Y(t) = \begin{cases} u^1 + \varepsilon^1(t), & t < \tau \\ u^2 + \varepsilon^2(t), & t \geq \tau \end{cases}$$

where u^1 and u^2 are constants,

$\varepsilon^1 \sim$ iid $N[0, \sigma_1^2]$,

$\varepsilon^2 \sim$ iid $N[0, \sigma_2^2]$, ε^1 is independent of ε^2.

The model used in Lahiri and Wang (1994) for the possibly differentiated series is

$$Y(t) = \begin{cases} u^1 + \varepsilon^1(t), & t < \tau \\ u^2 + \varepsilon^2(t), & t \geq \tau \end{cases}$$

where u^1 and u^2 are constants,

$\varepsilon^1(t) = \omega_1(t) + \phi_{11}\omega_1(t\text{-}1) + \ldots + \phi_{r1}\omega_1(t\text{-}r)$, $\omega_1 \sim$ iid $N[0, \sigma_1^2]$,

$\varepsilon^2(t) = \omega_2(t) + \phi_{12}\omega_2(t\text{-}1) + \ldots + \phi_{r2}\omega_2(t\text{-}r)$, $\omega_2 \sim$ iid $N[0, \sigma_2^2]$.

A common model for a stationary series is that the constant mean depends on the state (recession or expansion). If the differentiated series, $X(t)$-$X(t$-1), has a constant mean, then the expected values of the undifferentiated series, $X(t)$, are modeled as linear functions

$$\begin{cases} u^{01} + u^1 \cdot t, & t < \tau \\ u^{02} + u^2 \cdot t, & t \geq \tau \end{cases}$$

where $t=\{1, 2, ...\}$.

The stochastic term of the differentiated series is often assumed to be independent and normally distributed. That implies that the undifferentiated series has the following stochastic term

$$\begin{cases} \xi^1(t), & t < \tau \\ \xi^2(t), & t \geq \tau \end{cases}$$

$\xi^1 \sim$ iid N[0, $\sigma_1^2/2$],

$\xi^2 \sim$ iid N[0, $\sigma_2^2/2$], ξ^1 is independent of ξ^2.

The differentiation is made in order to separate the trend from the cyclic movements. Canova (1998) discusses trend adjustment and evaluates the effect of trend adjustment using several different approaches, among them first order differentiating. One conclusion from his study is that for some methods for trend adjustment, e.g. a polynomial with structural change, the resulting turning points agree with official data (The National Bureau of Economic Research, USA), whereas for example linear trend adjustment does not result in turning points that correspond to official data. In another paper Canova (1999) discusses that previous research has pointed out that trends vary over time and may interact in a nontrivial way with the cyclical component and therefore are difficult to isolate. Canova (1999) compares twelve methods for trend adjustment methods and two dating rules and the general conclusion is that statements concerning the turning points are not independent of the statistical assumptions needed to extract trends. For the data set that is investigated, first order differentiating resulted in a false alarm rate between 25% and 100% (depends on type of turning point and dating rule) and a missed-signal-rate between 28% and 100%.

Surveillance is here made in order to detect the next turning point. This means that the part of the series X that is monitored contains one turning point at most. Thus no separation of trend from cycle is made.

The model used in this report for an observation of the time series at time t is

$$X(t) = \mu(t) + \varepsilon(t) \qquad (1)$$

where $\mu(t) \in \wp$, $\wp$ is the family of all unimodal functions
and $\varepsilon(t) \sim$ iid N(0; σ^2), where σ^2 is assumed to be known.
Without loss of generality $\sigma^2 = 1$ is used in this investigation.

The major difference between model (1) and the models discussed above, is that the only knowledge about μ that is used in model (1), is the aspects of monotonicity and unimodality, which follows from the definition of a turning point in (2). In the turning point detection situation the aim at decision time s, is to discriminate between the two events $C(s)=\{\tau \leq s\}$ and $D(s)=\{\tau > s\}$, where τ is the unknown time of the turning point i.e.

$$C(s): \mu(t) \leq ... \leq \mu(\tau\text{-}1) \text{ and } \mu(\tau\text{-}1) \geq \mu(\tau) \geq ... \geq \mu(s) \qquad (2)$$

where $\tau \in \{1, 2, ..., s\}$
and at least one inequality is strict in the second part,

$$D(s): \mu(t) \leq ... \leq \mu(s).$$

The opposite case (i.e. detecting a trough) is handled analogously.

2.2 Method

The likelihood ratio method of surveillance has several optimality properties, see Frisén and de Maré (1991). Here we use the maximum likelihood ratio. The maximum likelihood ratio at time s is generally given by

$$\mathrm{MLR}(s) = \frac{\max f\{x_s|C\}}{\max f\{x_s|D\}} > k_s$$

where $C=\{\tau \leq s\}$ and $D=\{\tau > s\}$
and $f(x_s)$ is the likelihood function.

The event $D=\{\tau > s\}$ implies

$$\mu(t) \leq \mu(t+1),\ t \geq 1. \tag{3a}$$

The event $C=\{\tau \leq s\}$ is

$$C = \{\tau \leq s\} = \{\tau=1,\ \tau=2,\ ...,\ \tau=s\} = \{C_1, C_2, ..., C_s\},$$

where Cj implies

$$\mu(1) \leq ... \leq \mu(j\text{-}1) \text{ and } \mu(j\text{-}1) \geq \mu(j) \geq ..... \tag{3b}$$

Thus, in the MLR(s) we have

$$\max f\{x_s|D\} = f\{x_s|\hat{\mu}^D\}, \tag{4a}$$

where $\hat{\mu}^D$: $\max_{\mu \in F^D} f(x_s|\mu)$,

F^D is the family of μ such that $\mu(1) \leq \mu(2) \leq ... \leq \mu(s)$.

That is, $\hat{\mu}^D$ is the maximum likelihood estimator of μ under the monotonicity restriction D, see Robertson et al. (1988).

$$\max f\{x_s|C\} = \tag{4b}$$

$$\sum_{j=1}^{s}\left(\frac{P(\tau=j)}{P(\tau \leq s)}\right)\cdot\left(\max f\{x_s|Cj\}\right) =$$

$$\sum_{j=1}^{s}\left(\frac{P(\tau=j)}{P(\tau \leq s)}\right)\cdot\left(f\{x_s|\hat{\mu}^{Cj}\}\right),$$

where $\hat{\mu}^{Cj}$: $\max_{\mu \in F^{Cj}} f(x_s|\mu)$,

F^{Cj} is the family of μ such that $\mu(t) \leq ... \leq \mu(j\text{-}1)$ and $\mu(j\text{-}1) \geq \mu(j) \geq ....$

That is $\hat{\mu}^{Cj}$, $j \in \{1, 2, ..., s\}$, is the maximum likelihood estimator of μ under the monotonicity restriction Cj, given by Frisén (1986).

Thus, μ is estimated using a non-parametric method and the maximum likelihood ratio is given by

$$\mathrm{MLR}(s) = \frac{\max\left[f(x_s|C)\right]}{\max\left[f(x_s|D)\right]} = \frac{\sum_{j=1}^{s}\left(\frac{P(\tau=j)}{P(\tau\le s)} f(\mathrm{x_s}|\hat{\mu}^{Cj})\right)}{f(\mathrm{x_s}|\hat{\mu}^{D})} \tag{5}$$

where $f(x_s|\mu=\mu') = \left(\sqrt{2\pi}\right)^{-s} \exp\left(-\sum_{t=1}^{s}(x(t)-\mu'(t))^2\right)$

and $\hat{\mu}^D$ and $\hat{\mu}$ are defined in (4a) and (4b).

MLR(s) depends on the distribution of τ (the turning point time) in such a way that the likelihood function, conditional on C, is a weighted sum. In this report the Shiryaev-Roberts approach with a non-informative prior (equal weights) is used. This corresponds to the limiting distribution of the likelihood when the intensity, $\nu_t = P(\tau = t|\tau \ge t)$, tends to zero. Hereafter the method based on the maximum likelihood ratio, using the Shiryaev-Roberts approach, is referred to as the MSR method. Using the model specified in (1), the alarm statistic at time s can be written

$$\mathrm{MSR}(s) = \sum_{\mathrm{j=1}}^{\mathrm{s}} \exp\left\{\left(-\frac{1}{2}\right)\left(\sum_{\mathrm{t=1}}^{\mathrm{s}}(x(t)-\hat{\mu}^{\mathrm{Cj}}(\mathrm{t}))^2 - \sum_{\mathrm{t=1}}^{\mathrm{s}}(x(t)-\hat{\mu}^{\mathrm{D}}(t))^2\right)\right\}.$$

The Shiryaev-Roberts method approximately satisfies the optimality criterion of Shiryaev (1963) for small values of the intensity. Frisén and Wessman (1999) demonstrated that, using the Shiryaev-Roberts method as an approximation of the LR method for detecting a change from one level to another, is quite a good approximation, even for as large intensities as 0.20.

The time of the alarm, t_A, for the MSR method is defined as

$$t_A = \min[t: MSR(t) > k], \tag{6}$$

where the alarm limit, k, is a constant.

3 PERFORMANCE OF THE MSR METHOD

3.1 Theoretical study on the impact of the last observation and the post-peak slope

The delay of an alarm depends on both the method of surveillance and several process characteristics. In turning point detection, one characteristic is the post-peak slope. We investigate the influence of the post-peak slope, starting with the influence of the expected value of the last observation, $\mu(s)$. The alarm statistic at time s is written

$$\mathrm{MSR}(s) = \sum_{j=1}^{s} \exp\left(-\frac{1}{2}\left[Q_s^{Cj} - Q_s^{D}\right]\right) \tag{7}$$

where $Q_s^{Cj} = \sum_{t=1}^{s} (x(t) - \hat{\mu}_s^{Cj}(t))^2$

and $Q_s^D = \sum_{t=1}^{s} (x(t) - \hat{\mu}_s^D(t))^2$.

For example the alarm statistic for the decision time s=2 is

$$\text{MSR}(2) = \exp\left(-\frac{1}{2}\left[Q_2^{C1} - Q_2^D\right]\right) + \exp\left(-\frac{1}{2}\left[Q_2^{C2} - Q_2^D\right]\right) =$$

$$\exp\left(-\frac{(x(1) - \hat{\mu}_2^{C1}(1))^2}{2} - \frac{(x(2) - \hat{\mu}_2^{C1}(2))^2}{2}\right) *$$

$$\exp\left(\frac{(x(1) - \hat{\mu}_2^{D}(1))^2}{2} + \frac{(x(2) - \hat{\mu}_2^{D}(2))^2}{2}\right) +$$

$$\exp\left(-\frac{(x(1) - \hat{\mu}_2^{C2}(1))^2}{2} - \frac{(x(2) - \hat{\mu}_2^{C2}(2))^2}{2}\right) *$$

$$\exp\left(\frac{(x(1) - \hat{\mu}_2^{D}(1))^2}{2} + \frac{(x(2) - \hat{\mu}_2^{D}(2))^2}{2}\right)$$

The alarm statistic *MSR*(2) includes three quadratic deviations, namely Q_2^{C1}, Q_2^{C2} and Q_2^D. These quadratic deviations depend on the intra-relation of the observations $\{x(0), x(1), x(2)\}$. In the general case the alarm statistic at time s is a function of s+1 quadratic deviations, $\{Q_s^{C1}, ..., Q_s^{Cs}\}$ and Q_s. Their dependencies on the expected value of the last observation, $\mu(s)$, are expressed in the following lemmas, with proofs in Appendix A.

Lemma 1.1: *Q_s^D is decreasing stochastically as $\mu(s)$ increases.*
Lemma 1.2: *$Q_s^{C1}, Q_s^{C2}, ..., Q_s^{Cs-1}$ is increasing stochastically as $\mu(s)$ increases.*
Lemma 1.3: *Q_s^{Cs} is independent of $\mu(s)$.*

From Lemma (1.1)-(1.3) the following conclusion can be made regarding how the alarm statistic depends on the expected value of the last observation:

Theorem 1: *The alarm statistic MSR(s) is decreasing stochastically as $\mu(s)$ increases.*

Proof: *From Lemma (1.1)-(1-3) it follows that*

$(Q_s^{Cj} - Q_s^D)$, $j \in \{1, 2, ..., s\}$, increases stochastically as $\mu(s)$ increases.

Therefore, the alarm statistic $MSR(s) = \sum_{j=1}^{s} \exp\left(-\frac{1}{2}\left[Q_s^{Cj} - Q_s^{D}\right]\right)$ *decreases stochastically as* $\mu(s)$ *increases.* ■

The cumulative alarm probability at s, $P(t_A \leq s)$, depends on the expected value of the last observation, $\mu(s)$. This dependency is expressed in Corollary 1:

Corollary 1: *$P(t_A \leq s)$ is decreasing with $\mu(s)$.*

Proof: *From Theorem 1 it follows that* $P(MSR(s) > k \mid \bigcap_{j<s} MSR(j) < k)$ *decreases stochastically as* $\mu(s)$ *increases, since* $P\left(\bigcap_{j<s} MSR(j) < k\right)$ *is independent of* $\mu(s)$*. Thus* $P(t_A = s \mid t_A \geq s)$ *decreases as* $\mu(s)$ *increases. The cumulative alarm probability at time s is*

$$P(t_A \leq s) =$$
$$P(t_A \leq s-1) + P(t_A = s \mid t_A \geq s) \cdot P(t_A \geq s) =$$
$$P(t_A \leq s-1) + P(t_A = s \mid t_A \geq s) \cdot (1 - P(t_A \leq s-1)).$$

$P(t_A \leq s)$ decreases as $\mu(s)$ increases, since $P(t_A \leq s-1)$ is independent of $\mu(s)$ and $P(t_A = s \mid t_A \geq s)$ decreases as $\mu(s)$ increases. ■

After having examined the influence of the last observation, we use these results to investigate the influence of the post-peak slope. We express these results for two unimodal vectors, μ^A and μ^B, such that

$\mu^A(1) <...< \mu^A(\tau\text{-}1)$ and $\mu^A(\tau\text{-}1) > \mu^A(\tau) >...,$
and $\mu^B(1) <...< \mu^B(\tau\text{-}1)$ and $\mu^B(\tau\text{-}1) > \mu^B(\tau) >...,$

where $\mu^A(t) = \mu^B(t), \forall\, t \leq \tau\text{-}1$
and $\{\mu^A(t\text{-}1) - \mu^A(t)\} < \{\mu^B(t\text{-}1) - \mu^B(t)\}, \forall\, t \geq \tau.$

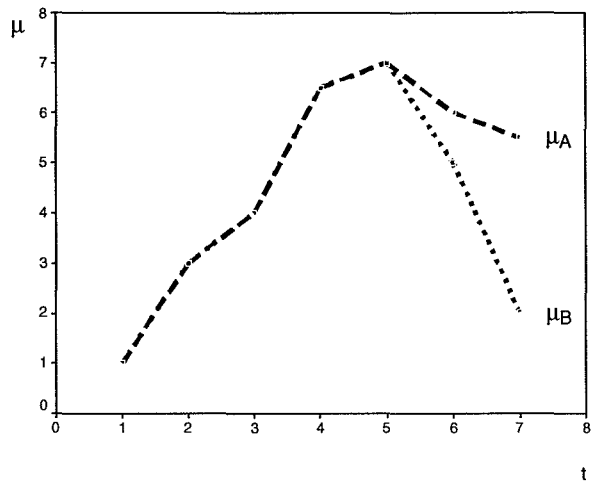

Figure 1a: *Illustration of* μ^B (. . .) *and* μ^A(▬ ▬ ▬), *for* $\tau = 6$.

To derive results regarding the distribution function for the time of alarm, three lemmas are needed, Lemma (2.1.1) - (2.2), which are given in Appendix B. The result regarding the distribution function for the time of alarm is presented in Theorem 2.

Theorem 2: *For two unimodal vectors* μ^A *and* μ^B, *as defined above, we have that* $P^B(t_A \leq s) > P^A(t_A \leq s)$, $\forall s \geq \tau$.

Proof: Given in Appendix C. ■

The result from Theorem 2 is used to derive the result regarding the conditional expected delay, presented in Theorem 3.

Theorem 3: *For two unimodal vectors* μ^A *and* μ^B, *as defined above, we have that* $E[(t_A - \tau_0) | t_A \geq \tau_0, \mu = \mu^B)] < E[(t_A - \tau_0) | t_A \geq \tau_0, \mu = \mu^A)]$.

Proof: *Since* t_A *is a discrete variable, defined only for non-negative integers, the expected value can be expressed as*

$$E[t_A] =$$
$$P(t_A = 1) + 2 \cdot P(t_A = 2) + 3 \cdot P(t_A = 3) + \ldots =$$
$$P(t_A > 0) + P(t_A > 1) + P(t_A > 2) + P(t_A > 3) + \ldots =$$
$$\sum_{t=0}^{\infty} P(t_A > t) =$$
$$\sum_{t=0}^{\infty} (1 - P(t_A \leq t))$$

The difference between the expected values is

$$E[t_A | \mu = \mu^B, \tau = \tau_0] - E[t_A | \mu = \mu^A, \tau = \tau_0] =$$
$$\sum_{t=0}^{\infty} P^A(t_A \leq t) - P^B(t_A \leq t) < 0,$$

$$since \begin{cases} P^A(t_A \le t) = P^B(t_A \le t), & t < \tau_0 \\ P^A(t_A \le t) < P^B(t_A \le t), & t \ge \tau_0 \end{cases}$$

The difference between the conditional expected values is

$$E[(t_A - \tau_0) \mid t_A \ge \tau_0, \mu = \mu^B] - E[(t_A - \tau_0) \mid t_A \ge \tau_0, \mu = \mu^A] =$$

$$\sum_{j=0}^{\infty} \frac{P^A((t_A - \tau_0) \le j)}{P^A(t_A \ge \tau_0)} - \frac{P^B((t_A - \tau_0) \le j)}{P^B(t_A \ge \tau_0)} < 0,$$

since $P^A(t_A < \tau_0) = P^B(t_A < \tau_0)$
and $P^A((t_A - \tau_0) \le j) < P^B((t_A - \tau_0) \le j)$ *for* $j \ge 0$. ■

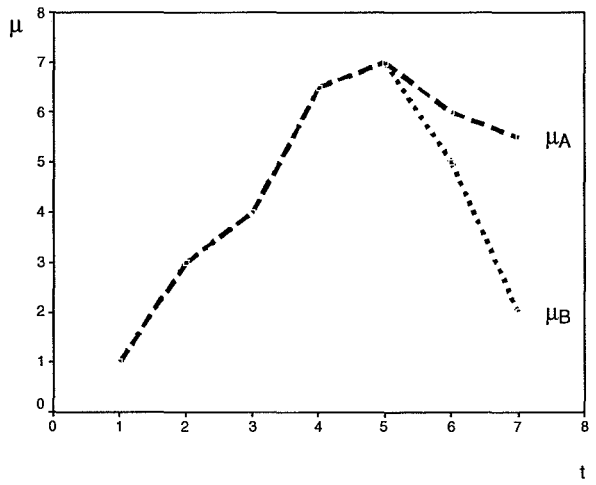

Figure 1b: *The conditional expected delay of an alarm (CED) is shorter for a vector* μ^B (. . .) *compared to a vector* μ^A (▬ ▬). *Illustration for* $\tau = 6$.

Thus, regardless of the parametric shape of the post-peak slope, the MSR method has the property that the conditional expected delay is shorter for a post-peak slope that is steeper at every time point.

3.2 A simulation study regarding the impact of the shape of the turning point

3.2.1 Models for the four simulated turning points

In Section 3.1 it was showed that the expected delay was shorter for a peak with a steeper post-peak slope. A simulation study is made in order to investigate some cases, where two factors, the symmetry and smoothness of the vector μ, influence the performance of the alarm statistic. The turning point is modeled using four different models, presented below. The first case (hereafter denoted the reference case) is when the turning point is modeled as:

$$X(t) = \mu_R(t) + \varepsilon(t) \quad (8)$$

where $\mu_R(t) = \beta_0 + \beta_R \cdot t - 2 \cdot D_1 \cdot \beta_R \cdot (t - \tau + 1)$

$t \in \{1, 2, ...\}$

$\beta_R = 0.26$

$\tau = 36$

$$D_1 = \begin{cases} 0, t < \tau \\ 1, t \geq \tau \end{cases}$$

$\varepsilon(t) \sim$ iid N[0; 1].

The value $\beta_R = 0.26$ is estimated from data on Swedish Industrial Production index, whereas β_0 is of no importance to the surveillance method, since the data can easily be transformed by $x(t)$-$x(0)$.

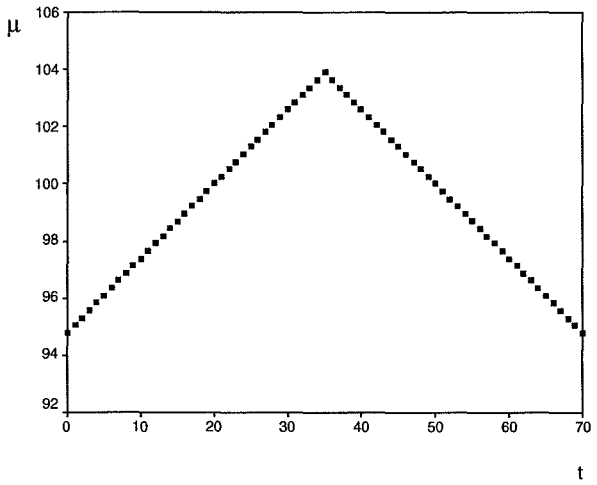

Figure 2: *The expected value as a function of time. Reference case, τ=36.*

Andersson (1999) evaluated the performance of the MSR method for the reference case by a simulation study. The other three cases investigated here are

μ is modeled as being symmetric with a non-smooth curve (case A)

μ is modeled as being symmetric with a smooth curve (case B)

μ is modeled as being non-symmetric with a smooth curve (case C)

In the reference case, the model used for μ was a linear trend with a symmetrical peak. The absolute growth of μ was constant. In the other three cases, denoted A, B and C, the vectorμ has a non-constant growth and a plateau at the peak.

For case A (symmetric and non-smoothed), the vector $\mu_A(t)$ is based on the Swedish Industrial Production index, which has been deseasonalised and smoothed. The vector μ_A is not a linear function. However, the average slope of μ_A (both pre-peak and post-peak) is the same as for the reference case, i.e. for the pre-peak slope we have

$$\mu_A(t) = \beta_R \cdot t + a(t), t < \tau \quad (9)$$

where $\sum_{j=1}^{\tau-1} a(j) = 0$

$\tau = 36$

Var[a] = 0.32.

The post-peak slope is modeled correspondingly. Thus μ_A has the same slope as the reference case (β_R) but a non-constant growth (modeled by a in (9)). This characteristic (that μ_A has a non-constant growth) is hereafter called non-smoothness. The non-smoothness in case A results in a higher variance for case A, than for the reference case, i.e. $\mathrm{Var}[X_R(t)]=\sigma^2=1$ and $\mathrm{Var}[X_A(t)]=\mathrm{Var}[a]+\sigma^2 = 0.32 + 1$. In the reference case, the constant growth of the slope results in a sharp peak. This can be compared to case A (Figure 3), where the peak is non-sharp (that is, the growth rate of μ_A is low just around the peak).

For case B (symmetric and smoothed) the vector $\mu_B(t)$ is modeled using trigonometric functions, which results in a smooth curve with a non-sharp peak. The vector $\mu_B(t)$ is not a linear function. However, the slope of μ_B (both pre-peak and post-peak) for case B is the same as for the reference case, except at the top. The model used is

$$\mu_B(t) = \beta_0 + \beta_1 \cos[\gamma_1\pi(t+2)] + \beta_2 \sin[\gamma_1\pi(t+2)]. \tag{10}$$

The peak in case B is non-sharp, compared to the peak in the reference case (Figure 3).

For case C (non-symmetric and smoothed) the pre-peak vector μ_C is modeled according to the same model used for case B. The pre-peak slope of μ_C is the same as for the reference case, except at the top. In order to achieve a non-symmetrical peak where the post-peak slope, $\mu_C(t)$ for $t \geq \tau$, is steeper, the modeling is made using different trigonometric functions compared to the pre-peak slope. The model used for $\mu_C(t)$ is

$$\mu_C(t) = \begin{cases} \beta_0 + \beta_1 \cos[\gamma_1\pi(t+2)] + \beta_2 \sin[\gamma_1\pi(t+2)], & t < \tau \\ \beta_3 + \beta_4 \cos[\gamma_2\pi(t-7)] + \beta_5 \sin[\gamma_2\pi(t-7)], & t \geq \tau. \end{cases} \tag{11}$$

The peak of case C is non-sharp (Figure 3), compared to the reference case. The post-peak slopes of μ_R and μ_C are practically the same for the period $36 \leq t \leq 45$. When comparing case B and C (Figure 3), the pre-peak slopes are identical and case C has a steeper post-peak slope.

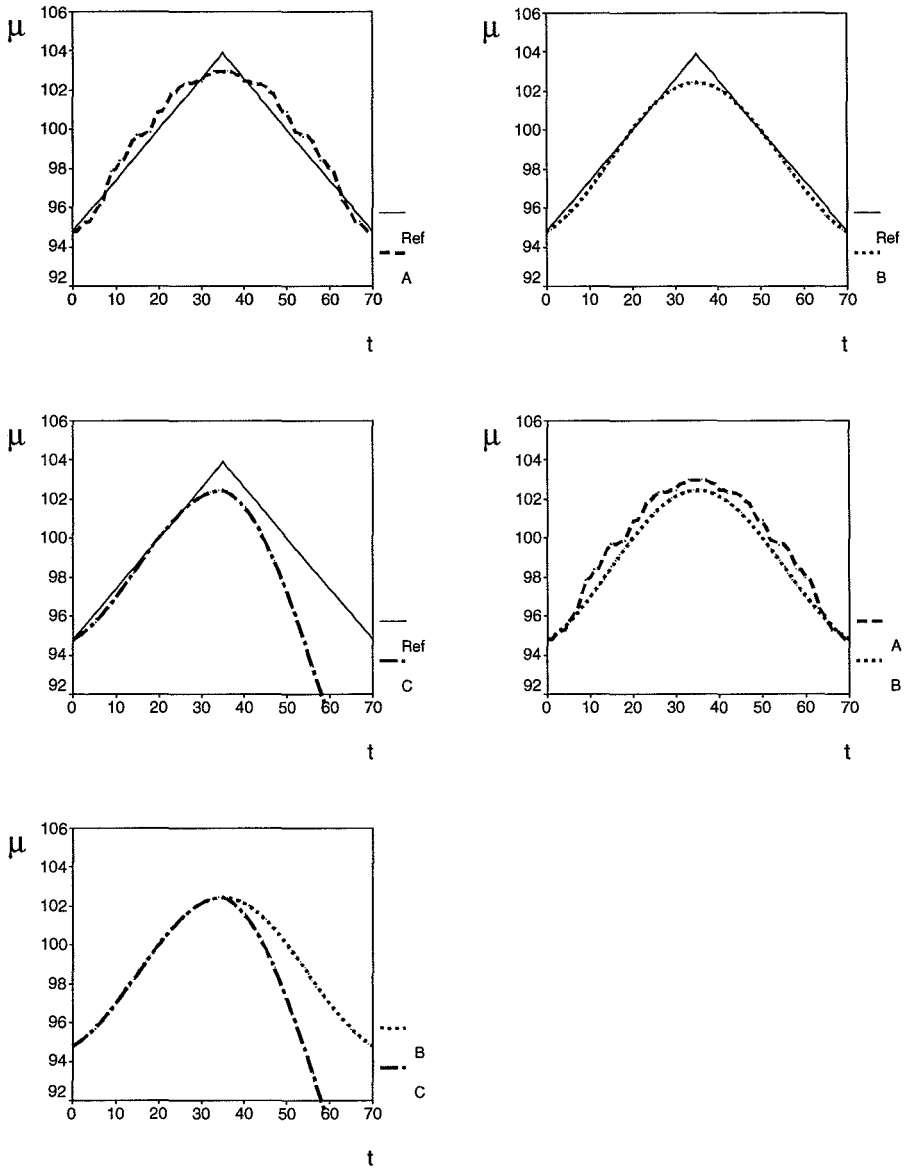

Figure 3: *The expected value as a function of time.*
Case A and reference case (left, top), case B and reference case (right, top), case C and reference case (left, middle), case A and B (right, middle), case B and C (left, bottom).

3.2.2 The probability of a false alarm

Generally in surveillance, the way in which false alarms for turns are controlled is important. In the general theory and practice of surveillance, the most common way is to control the ARL^0, (the Average Run Length to the first alarm if the process does not have any turn). Hawkins (1992), Gan (1993) and Andersson (1999) suggest that the control is made by a statistic similar to the ARL^0, namely the MRL^0, which is the median run length, which has several advantages, such as easier interpretations for the skewed distributions and much shorter computer time for calculations. The approach used in much theoretical work e.g. Shiryaev (1963) and Frisén and de Maré (1991) and for which optimality theorems are available, is a control of the probability of false

alarm, $P(t_A<\tau)$. The alarm limit is determined to yield a fix false alarm probability. Neftci (1982) and Lahiri and Wang (1994) use this criterion for alarms for turning points of business cycles. In this paper the probability of a false alarm, $PFA_{36}=P(t_A<\tau|\tau=36)$, is set equal for all four investigated cases, $\hat{PFA}_{36}=0.31$. The standard deviation of an estimate of a value v is hereafter denoted sd[$\hat{v}$] and so sd[$\hat{PFA}_{36}$] =0.0060. The distribution functions of the false alarm for all cases are shown in Figure 4. The pre-peak slopes of case B and C are the same, i.e. $\mu_B(t) = \mu_C(t)$, $t < \tau$. Therefore, only three functions are presented in Figure 4.

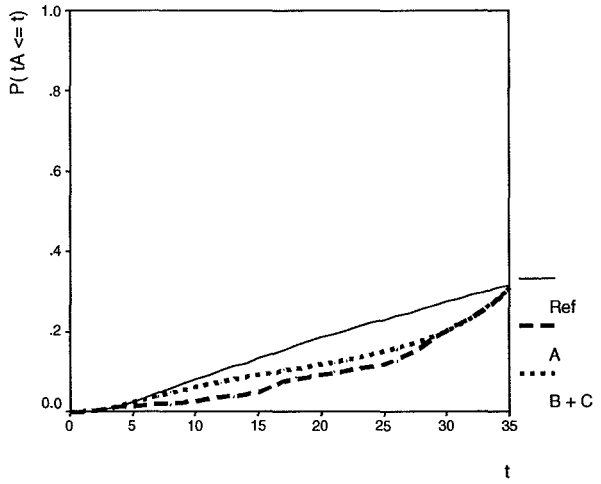

Figure 4: *The distribution functions for the false alarm: Reference (linear), A (symmetric, non-smooth), B (symmetric, smooth), C (non-symmetric, smooth).*

The results in Figure 4 above illustrate that the distribution function of the false alarm depends on the smoothness and shape of the pre-peak μ-vector. The non-smoothness of $\mu_A(t)$ is reflected in the distribution function of the false alarms for case A. For all three cases A, B and C, the peak is non-sharp (i.e. the growth rate of μ is decreasing continuously around the peak). As a result, the false alarm rate for cases A, B and C is increasing just before the peak, which will have an impact on the behavior of MSR after the turning point.

3.2.3 The delay time

The delay time is measured using both CED(τ_0) (see Section 3.1) and the conditional median delay, CMD(τ_0) = M[$t_A-\tau|t_A\geq\tau=\tau_0$], where M[•] is the median. The variance of the delay time differs between the four investigated peaks, since it depends on the post-peak slope of the μ-vector. The number of replicates used to estimate CED(36) and CMD(36) are 13 000 (reference case), 33 000 (case A), 30 000 (case B) and 15 000 (case C). The number of replicates result in sd[$\hat{\text{CED}}$(36)]=0.018 for all cases.

The distribution functions of t_A (time of alarm) for the case when τ= 36 are shown in Figure 5. The results concerning the delay are summarized in Table 1.

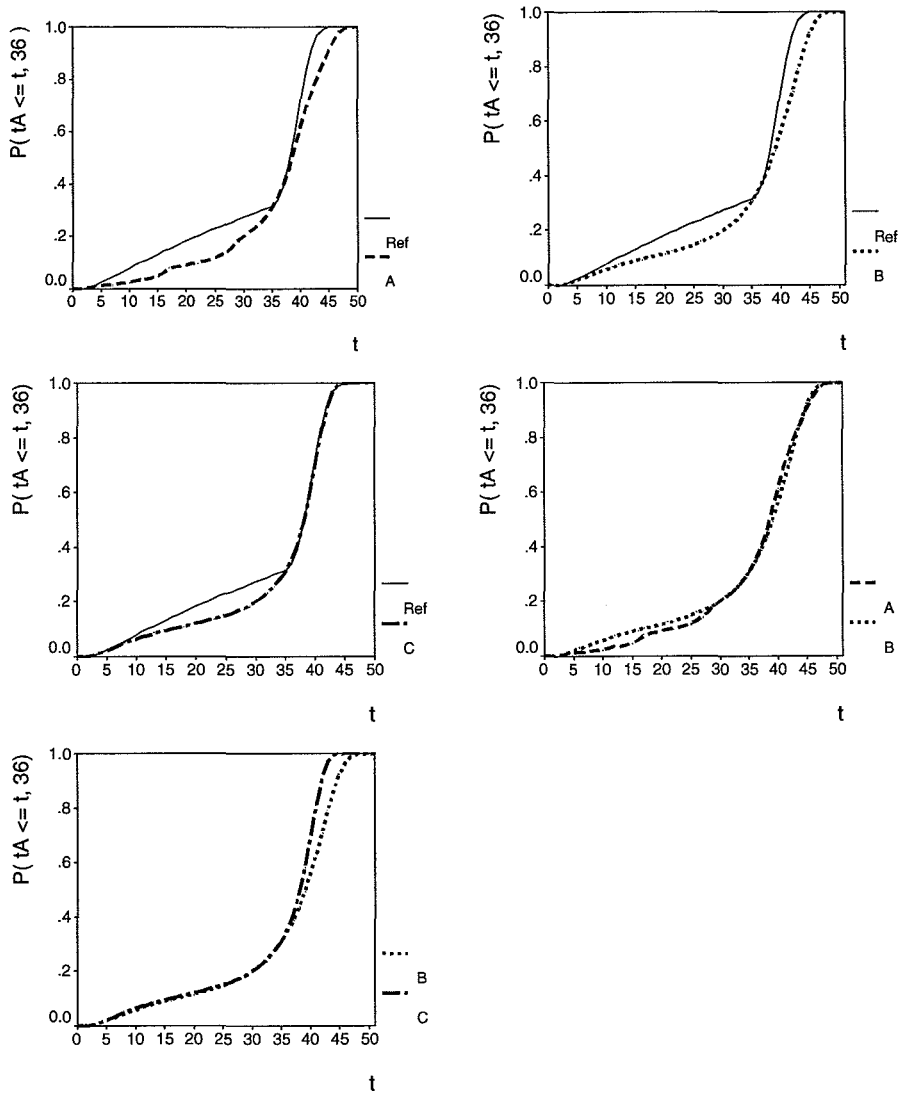

Figure 5: *The distribution of the time of alarm (false and motivated), when τ=36. Left, top: Reference (linear) and A (symmetric, non-smooth). Right, top: Reference (linear) and B (symmetric, smooth). Left, middle: Reference (linear) and C (non-symmetric, smooth). Right, middle: A (symmetric, non-smooth) and B (symmetric, smooth). Left, bottom: B (symmetric, smooth) and C (non-symmetric, smooth).*

Table 1: *Conditional expected delay and conditional median delay, (sd[•] in parenthesis).*

Case	$\hat{CED}(36)$	$\hat{CMD}(36)$
Reference	4.05 (0.018)	3.51 (0.022*)
Symmetric and non-smooth (A)	5.35 (0.018)	4.54 (0.023*)
Symmetric and smooth (B)	5.66 (0.018)	5.29 (0.030*)
Non-symmetric and smooth (C)	4.06 (0.018)	3.68 (0.024*)

*: based on a simulation study

The results in Figure 5 and Table 1 above illustrate that the conditional expected delay and the conditional median delay depend on the shape of μ around the turning point. On one hand a non-sharp peak (case A, B and C) results in an increase in the alarm statistic just before the turning point. Thereby only a small increase in the alarm statistic at $s = \tau$ is needed to yield an alarm. The result is a decrease in CED and CMD. On the other hand, the characteristics of the peak just after the turning point (sharpness) will affect the alarm statistic in the opposite direction: CED and CMD are longer as a result of the non-sharp peak. An expansion (or recession) phase that is non-smooth has a similar effect as an increased variance. In the simulations, the non-smooth peak has a higher alarm limit, so as to deal with the higher tendency to a false alarm. It is shown that, after adjustment for false alarms, the non-smoothness does not have a large effect. The longest delay is for peak B (symmetrical, non-sharp and smooth), but the difference in delay to peak A (symmetrical, non-sharp and non-smooth) is not large.

3.2.4 The probability of successful detection

The measures of performance presented in Section 3.2.3 are CED(36) and CMD(36). The performance of a system of surveillance can be evaluated using other measures, see Frisén (1992). The measures of evaluation depend on the application. In applications where early warnings are important, such as business cycles, the delay time is a relevant measure. In other applications of turning point detection, for example natural family planning, see Royston (1991), the success of the surveillance depends on the ability to detect the peak within a certain limited time interval, d. Hence the probability of successful detection, Frisén (1992), is a relevant measure of the performance. The probability of successful detection is defined as

$$\mathrm{PSD}(\tau, d) = P\left(t_A - \tau < d \middle| t_A \geq \tau, \tau = \tau_0\right) \qquad (12)$$

The probability of successful detection for d=3, d =5 and d =11 is given in Table 2.

Table 2: *Probability of successful detection (sd[•] in parenthesis).*

	$\hat{PSD}(\tau = 36, d)$		
Case	$d = 3$	$d = 5$	$d = 11$
Reference	0.23 (0.0037)	0.59 (0.0043)	1.00 (0.0001)
Symmetric and non-smooth (A)	0.21 (0.0022)	0.44 (0.0027)	0.93 (0.0014)
Symmetric and smooth (B)	0.18 (0.0022)	0.36 (0.0028)	0.95 (0.0012)
Non-symmetric and smooth (C)	0.25 (0.0036)	0.55 (0.0041)	1.00 (--------)

Figure 6 shows PSD(τ=*36*, *d*) as a function of *d*.

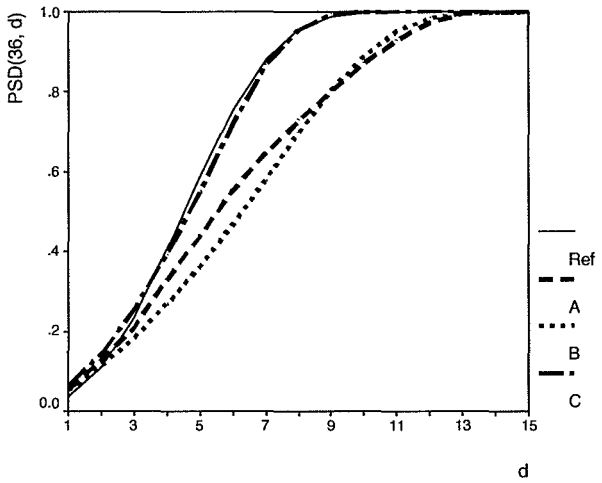

Figure 6: *The probability of successful detection, when $\tau = 36$. Reference (linear), A (symmetric, non-smooth), B (symmetric, smooth), C (non-symmetric, smooth).*

The PSD measure is a complement to the CED and CMD. The results in Figure 6 and Table 2 above illustrate that ranking between the four cases of turning points is independent of the measurement time: at all times presented in Table 2, the reference case and case C have the shortest time until detection. Case B (symmetric and smooth) is the most difficult to find. A non-sharp peak, where the post-peak slope is steeper, is detected equally fast as a sharp, symmetrical peak.

The probability of successful detection, PSD(τ, *d*), is important when only alarms within a limited time after the change is of use. For a turning point situation where the actual peak is preceded by a plateau, also (false) alarms just before the change might be useful. In order to evaluate the detection probability around the peak, the following probability is used

$$\text{absPSD}(\tau_0, d) = P\left(|t_A - \tau_0| < d\right). \tag{13}$$

Bojdecki (1979) gives the solution to a maximization of absPSD(τ_0, d), where d is a constant integer.

The probability of successful detection within d=3 and d=5 time units from the peak is given in Table 3.

Table 3: *Probability of successful detection within an interval of absolute length (sd[•] in parenthesis).*

	$\widehat{\text{absPSD}}(36, d)$	
Case	$d = 3$	$d = 5$
Reference	0.18 (0.0027)	0.44 (0.0035)
Symmetric and non-smooth (A)	0.20 (0.0018)	0.39 (0.0022)
Symmetric and smooth (B)	0.18 (0.0018)	0.34 (0.0023)
Non-symmetric and smooth (C)	0.25 (0.0029)	0.52 (0.0034)

The function absPSD(τ, d) is presented in Figure 7 for τ=36, $d \leq 5$.

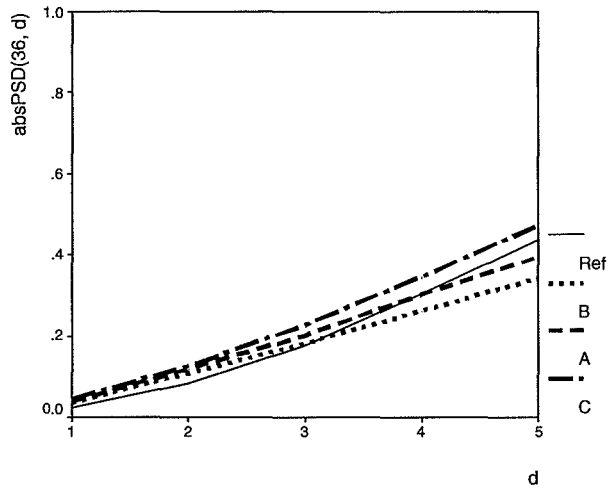

Figure 7: *The probability of successful detection within an interval of absolute length d.Reference (linear), A (symmetric, non-smooth), B (symmetric, smooth), C (non-symmetric, smooth).*

The results in Figure 7 and Table 3 above illustrate that the probability of detection depends on the shape of the μ-vector just around the peak. A non-sharp peak is equally easy to detect as a sharp peak, because of the many alarms just before the peak.

4 OUT-OF-SAMPLE PERFORMANCE FOR SWEDISH DATA

Official data over two series of the Swedish Import index, Figure 8, are used to illustrate the performance of the MSR method. Data are for the period January, 1990 to December 1994 and are adjusted for seasonality (source: Statistics Sweden).

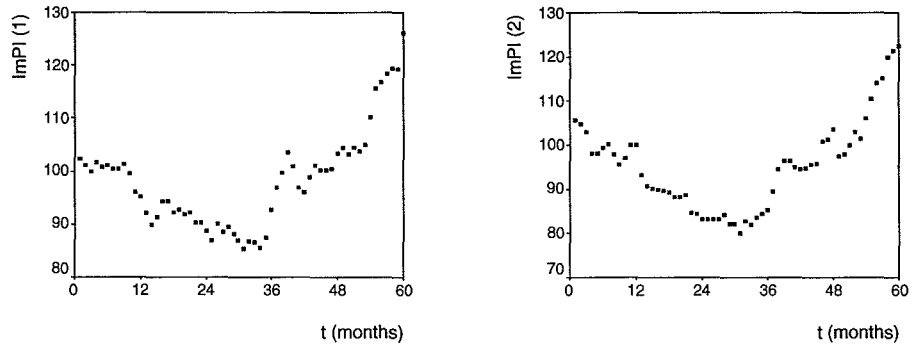

Figure 8: *Monthly observations (9001: 9412) on Swedish Import index (adjusted for seasonality), for two different areas of metal import.*

For the recession phase of both troughs the signal/noise ratio, β/σ, equals (0.26/1.00) and linear transformations, $x'(t) = x(t)/\hat{\sigma}$, suffice to standardize the observations before the MSR method is applied. Evaluation of a statistical method is sometimes made on a set of data that has not been used in the model-building process, in order to test the out-of-sample performance. The performance of a method of surveillance on a specific data set can be evaluated by reporting e.g. the delay, the false alarms or the Brier's probability score. The latter is defined as

$$\frac{1}{s}\sum_{i=1}^{s}(j(s)-p(s))^2,$$

where $j(s) = \begin{cases} 0, & \text{if recession} \\ 1, & \text{if expansion} \end{cases}$

and $p(s) = P(J(s)=0|x_s)$.

The measures are calculated under different assumptions regarding intensity (see Section 2.2), namely $\nu = \{0.01, 0.1\}$.

Table 4: *The MSR method, investigated for the Swedish Import index.*

	ImPI(1)	**ImPI(2)**
Period of time	1990M1: 1994M12	1990M1: 1994M12
Time of trough*	1992M7	1992M7
First time point after trough, τ	1992M8	1992M8
Time of alarm, t_A	1992M12	1993M1
Delay, t_A-τ	4	5
False alarm	No	No
Brier's probability score, τ~Geo(ν=0.01)	0.03	0.08
Brier's probability score, τ~Geo(ν=0.1)	0.26	0.17

*: Using the minimum value

In Table 4, different measures are presented, both traditional forecasting measures (Brier's probability score, see e.g. Koskinen and Öller (1998)) and measures used in surveillance. As already pointed out, the time of the alarm is important in surveillance and therefore measures which reflect the timeliness must be used in the evaluation. This is a drawback with Brier's probability score: since the deviations are squared, the same measure is given for false alarms as for delayed alarms. Since Brier's probability score does not take the order of the observations into account, it is not a suitable criterion in a surveillance situation, where the order of the observations holds a large amount of the relevant information. As is seen in Table 4, the different assumptions of the intensity have a major impact on Brier's probability score.

5 DISCUSSION AND CONCLUDING REMARKS

The influence of the shape of the turning point on the MSR method of surveillance is investigated. The theoretical investigation regarding the influence of the post-peak slope shows that, regardless of which function that is true for the post-peak slope, the conditional expected delay of an alarm decreases as the post-peak slope grows steeper, for the same pre-peak slope. That is, if we have two processes, A and B, where the pre-peak slopes are the same, but with post-peak slopes such that the expected difference between consecutive observations is larger for B at every time point, then B has a shorter conditional expected delay. In business cycle theory, as a rule the expansion phase lasts longer and is flatter than the recession phase Oppenländer (1997). For the situation where the peak is non-symmetrical with steeper post-peak slope the result regarding the conditional expected delay indicates that a

transition from expansion to recession is quicker detected than the opposite transition, if the same alarm limit is used.

In a surveillance system, the probability of false alarms should be kept on an acceptable level, but it must be born in mind that if the probability of false alarms is low, the probability of detection is also low. In some applications, a relatively high probability of false alarms could be preferred to the risk of missing a motivated alarm, whereas in other applications the situation can be the opposite. In this study the probability of false alarms is set to 0.31. The study shows that the distribution of the false alarms depends on the slope and smoothness (in relation to the standard deviation of the observations) of the pre-peak slope, i.e. the slope and smoothness of $\mu(t)$, $t<\tau$. For the situation where μ has a constant growth, the false alarm rate is approximately constant. The effect of a non-smooth μ-vector is a non-constant false alarm rate. A non-sharp peak (i.e. where the growth of μ is decreasing just around the peak) results in an increasing false alarm rate just before the peak. This is not a disadvantage.

Preferably, a method of surveillance should have a small false alarm probability and a short expected delay of motivated alarms. The effect of the shape of the μ-vector on the conditional expected delay is investigated for a fixed false alarm probability. The study shows that the conditional expected delay decreases as the post-peak slope grows steeper. The conditional expected delay is sensitive to the shape of the μ-vector just around the peak. The conditional expected delay is shorter for a sharp peak (i.e. where μ has constant growth) compared to a non-sharp peak (i.e. where the growth of μ is decreasing just around the peak). A non-smooth μ-vector has a similar effect as an increased variance namely an increased conditional expected delay.

In some situations it is vital to detect the turning points within a limited time interval, for example situations where the time for effective actions is limited. The detection probability can be evaluated using the probability of successful detection. If only alarms after the turning point are considered, this study shows that the probability of successful detection increases for a steeper post-peak slope and decreases for a non-sharp peak. For all four cases investigated the probability of detection within 10 time units exceeds 0.90. In natural family planning, the aim is to detect the interval around the turning point. The most fertile phase of the human menstrual cycle occurs in the interval two days before the peak to three days after the peak in the oestrogen hormone, see Royston (1991). In this situation alarms around the peak is of interest. The evaluation by probability of successful detection showed that the detection probability is larger for a sharp peak. But for detection around the peak, i.e. alarms around the peak, the study shows that a non-sharp peak is equally easy to detect as a sharp peak, because of the many alarms just before the peak.

The performance of the MSR method was investigated on two sets of real data, with a satisfactory result. The evaluation also showed that the Brier probability score is very sensitive to assumptions regarding the intensity.

The MSR method uses a non-informative prior for the time of the turning point. Often, see e.g. Koskinen and Öller (1998), the turning point time is assumed to be geometrically distributed with a constant intensity. If the distribution of the turning point time is used in a posterior probability with a fixed limit (or corresponding alarm-condition by the likelihood ratio), it makes the alarm statistic very sensitive to the intensity parameter. Therefore, a robust approach with a non-informative prior is used here.

In many applications of turning point detection, for example business cycle prediction using more than one leading indicator and natural family planning using indicators like temperature and hormone level, several turning point indicators can be used. The conclusions about the process are improved if information from all relevant indicators is used for making the decision, i.e. if multivariate surveillance is applied. In multivariate surveillance of more than one cyclical time series, several process characteristics effect the performance of the surveillance method. Examples of such process characteristics are i) the relation between the turning point times of the processes, ii) the slopes and standard deviation of the processes and iii) the interdependence between the processes, given their change points. Regarding the relation between the turning point times, the turning points can either occur at the same time (or with known lags) or independent of each other. For a situation where all processes change at the same time (or with a known time lag), the sufficient alarm statistic can be determined. Then it is possible to reduce the multivariate surveillance to a univariate one by a summary statistic. This was done by Wessman (1998). If the processes are monitored separately, a decision procedure for multivariate inference, for example the method of union-intersection, see Roy (1953), can be used to decide when to call an alarm. Another process characteristic in multivariate surveillance of cyclical processes is the signal-noise ratio (the slope in relation to the standard deviation) of each process. In order for a process to posses any predictive ability, the signal noise ratio must not be too small, i.e. there is a limit as to whether the addition of one more process will contribute to the detection probability of the surveillance system. Yet another important factor to consider is whether the processes, given their change points, are independent of each other. Both in the situation of leading economic indicators and in the situation of natural family planning, the different processes are likely to be dependent. The situation where several dependent processes shifted from one level to another, was investigated by Wessman (1999).

ACKNOWLEDGEMENTS

I would like to thank my supervisor, Professor Marianne Frisén, for her encouraging guidance and support during the work with this report. I would also like to thank my colleague Christian Sonesson for valuable help with the draft manuscript and all my other friends and colleagues at the Department of Statistics. This report is part of a project on statistical surveillance, which is supported by the Swedish Council for Research in the Humanities and Social Sciences.

REFERENCES

Andersson, E. (2001) Monitoring cyclical processes. A non-parametric approach. To appear in *Journal of Applied Statistics*.

Arteaga, C. and Ledolter, J. (1997) Control charts based on order-restricted tests. *Statistics & Probability Letters,* **32,** 1-10.

Barlow, R. E., Bartholomew, D. J., Bremer, J. M. and Brunk, H. D. (1972) *Statistical inference under order restrictions,* Wiley, London.

Birchenhall, C. R., Jessen, H., Osborn, D. R. and Simpson, P. (1999) Predicting US business-cycle regimes. *Journal of Business & Economic Statistics,* **17,** 313-323.

Bojdecki, T. (1979) Probability maximizing approach to optimal stopping and its application to a disorder problem. *Stochastics,* **3,** 61-71.

Canova, F. (1998) Detrending and business cycle facts. *Journal of Monetary Economics,* **41,** 475-512.

Canova, F. (1999) Does detrending matter for the determination of the reference cycle and the selection of turning points? *The Economic Journal,* **109,** 126-150.

Frisén, M. (1986) Unimodal regression. *The Statistician,* **35,** 479-485.

Frisén, M. (1992) Evaluations of Methods for Statistical Surveillance. *Statistics in Medicine,* **11,** 1489-1502.

Frisén, M. (1994) Statistical Surveillance of Business Cycles. *Research report* 1994:1 (Revised 2000), Department of Statistics, Göteborg University, Sweden.

Frisén, M. and de Maré, J. (1991) Optimal Surveillance. *Biometrika,* **78,** 271-80.

Frisén, M. and Wessman, P. (1999) Evaluations of likelihood ratio methods for surveillance. Differences and robustness. *Communications in Statistics. Simulations and Computations,* **28,** 597-622.

Gan, F. F. (1993) An optimal design of EWMA control charts based on median run length. *Journal of Statistical Computation and Simulation,* **45,** 169-184.

Hamilton, J. D. (1989) A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica,* **57,** 357-384.

Hawkins, D. L. (1992) Detecting shifts in functions of multivariate location and covariance parameters. *Journal of Statistical Planning and Inference,* **33,** 233-244.

Jun, D. B. and Joo, Y. J. (1993) Predicting turning points in business cycles by detection of slope changes in the leading composite index. *Journal of Forecasting,* **12,** 197-213.

Koskinen, L. and Öller, L.-E. (1998) A Classifying Procedure for Signalling Turning Points. *Workning Paper*, National institute of economic research, Stockholm, Sweden.

Lahiri, K. and Wang, J. G. (1994) Predicting cyclical turning points with a leading index in a Markov Switching model. *Journal of Forecasting,* **13,** 245-263.

Layton, A. P. (1996) Dating and predicting phase changes in the U.S. business cycle. *International Journal of Forecasting,* **12,** 417-428.

Neftci, S. (1982) Optimal prediction of cyclical downturns. *Journal of Economic Dynamics and Control,* **4,** 225-41.

Oppenländer, K. H. (1997) The phenomenon of business cycles. In *Business cycle indicators* (Ed, Oppenländer, K. H.) Avebury, England, pp. 3-23.

Robertson, T., Wright, F. T. and Dykstra, R. L. (1988) *Order Restricted Statistical Inference,* John Wiley & Sons Ltd.

Roy, S. N. (1953) On a Heuristic Method of Test Construction and Its Use in Multivariate Analysis. *Annals of Mathematical Statistics,* **24,** 220-238.

Royston, P. (1991) Identifying the fertile phase of the human menstrual cycle. *Statistics in Medicine,* **10,** 221-240.

Shiryaev, A. N. (1963) On optimum methods in quickest detection problems. *Theory of Probability and its Applications.,* **8,** 22-46.

Srivastava, M. S. and Wu, Y. (1993) Comparison of EWMA, CUSUM and Shiryayev-Roberts Procedures for Detecting a Shift in the Mean. *Annals of Statistics,* **21,** 645-670.

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.

Wessman, P. (1999) The surveillance of several processes with different change points. *Research report* 1999:2, Department of Statistics, Göteborg University, Sweden.

Wetherill, G. B. and Brown, D. W. (1991) *Statistical Process Control: Theory and Practice,* London Chapman and Hall.

Zarnowitz, V. and Moore, G. H. (1982) Sequential signals of recession and recovery. *Journal of Business,* **55,** 57-85.

Appendix A. Proof of Lemma (1.1) - (1.3), used in Theorem 1.

The following model is used

$X(s) = \mu(s) + \varepsilon(s)$, where $\varepsilon(s) \sim$ iid N(0; 1)

Thus $X(s)$ increases stochastically with $\mu(s)$.

Lemma 1.1: Q_s^D is decreasing stochastically with $\mu(s)$

Proof: It follows from the results in Barlow et al. (1972) that the estimation procedure for the vector μ_s^D can be described as first partitioning the time points $j \in \{1, 2, ..., s\}$ into $m(s)$ sets, $\{ L_1^D, L_2^D, ..., L_{m(s)}^D \}$, where

$L_1^D = \{1,...,k_1\}$

$L_2^D = \{k_1+1,...,k_2\}$

...

$L_{m(s)}^D = \{k_{m(s)-1}+1,...,s\}$.

Then, for $\forall j \in L_h^D$,

$$\hat{\mu}_s^D(j) = \frac{\sum_{i \in L_h^D} X(i)}{k_h - k_{h-1}},$$

where the partitioning ensures that $\hat{\mu}_s^D(q) < \hat{\mu}_s^D(r)$ if q$\in L_h^D$, r$\in L_{h'}^D$, $h<h'$.

The ML-estimator $\hat{\mu}_s^D(s)$ has the property that either

$\hat{\mu}_s^D(s) = X(s)$

or

$$\hat{\mu}_s^D(s) = \frac{\sum_{j=k_{m(s)}}^{s} X(j)}{s - k_{m(s)} + 1} = \overline{X}(k_{m(s)}, s),$$

where $k_{m(s)} \leq (s\text{-}1)$ and $X(s) < \overline{X}(k,s)$.

$k_{m(s)}$ is denoted k in the following.

At time s-1 we have $Q_{s-1}^D = \sum_{j=1}^{s-1} \left(X(j) - \hat{\mu}_{s-1}^D(j)\right)^2 = \sum_{j=1}^{s-1} Q_{s-1}^D(j)$

where $\hat{\mu}_{s-1}^D(j)$, $j \in \{1, 2, ..., s-1\}$, are the maximum likelihood estimates for time points j.

The dependence of Q_s^D on $X(s)$ can be shown through the maximum likelihood estimate of $\hat{\mu}_s^D(s)$:

$$\hat{\mu}_s^D(s) = X(s) \Rightarrow Q_s^D(s) = 0 \quad \Rightarrow Q_s^D = \sum_{j=1}^{s-1} Q_s^D(j) = \sum_{j=1}^{s-1} Q_{s-1}^D(j) = Q_{s-1}^D$$

$$\hat{\mu}_s^D(s) = \overline{X}(k;s),\ X(s) < \overline{X}(k,s) \Rightarrow Q_s^D = \sum_{j=1}^{s} Q_s^D(j) = \sum_{j=1}^{s-1} Q_s^D(j) + Q_s^D(s)$$

The inequality $Q_s^D > Q_{s-1}^D$ holds since

$\sum_{j=1}^{s-1} Q_s^D(j) > \sum_{j=1}^{s-1} Q_{s-1}^D(j)$ and $Q_s^D(s) > 0$.

The inequality $\sum_{j=1}^{s-1} Q_s^D(j) > \sum_{j=1}^{s-1} Q_{s-1}^D(j)$ follows since

$\sum_{j=1}^{s-1} Q_{s-1}^D(j)$ is the least square sum under the monotonicity restrictions and $\exists j \in \{1, ..., s\text{-}1\}$: $Q_s^D(j) \neq Q_{s-1}^D(j)$.

Thus $min[Q_s^D] = Q_{s-1}^D$ is obtained for $\hat{\mu}_s^D(s) = X(s)$.

For $\hat{\mu}_s^D(s) = \overline{X}(k;s)$ we have

$$Q_s^D = \sum_{j=1}^{k-1} \left(X(j) - \hat{\mu}_s^D(j)\right)^2 + \sum_{j=k}^{s-1} \left(X(j) - \overline{X}(k;s)\right)^2 + \left(X(s) - \overline{X}(k;s)\right)^2.$$

The first term, $\sum_{j=1}^{k-1} \left(X(j) - \hat{\mu}_s^D(j)\right)^2$, is independent of $X(s)$.

The mid-term can be written as

$$\sum_{j=k}^{s-1} \left(X(j) - \overline{X}(k;s)\right)^2 = \sum_{j=k}^{s-1} \left(X(j) - \frac{(s-k)\overline{X}(k;s-1)}{s-k+1} - \frac{X(s)}{s-k+1} \right)^2 = f(X(s))$$

For each fixed k

$\frac{\partial f(X(s))}{\partial X(s)} = c_1 (X(s) - \overline{X}(k;s-1))$, where c_1 is a constant.

For $X(s) < \overline{X}(k;s-1)$ we have $\dfrac{\partial f(X(s))}{\partial X(s)} < 0$, thus, $f(X(s))$ is increasing as $X(s)$ is decreasing.

The last term, $\left(X(s) - \overline{X}(k;s)\right)^2$, can be written as

$$\left(\frac{s-k}{s-k+1}\right)^2 \left(X(s) - \overline{X}(k;s-1)\right)^2.$$

The term $\left(X(s) - \overline{X}(k;s)\right)^2$ increases as $X(s)$ decreases, since $X(s) < \overline{X}(k;s-1)$.

Thus the quadratic deviation Q_s^D decreases with $X(s)$.
Since $X(s)$ increases stochastically with $\mu(s)$, it follows that Q_s^D decreases stochastically with $\mu(s)$.■

Lemma 1.2: Q_s^{C1}, Q_s^{C2},..., Q_s^{Cs-1} is increasing stochastically with $\mu(s)$

Proof: The proof of Lemma 1.2 is made correspondingly to the proof of Lemma 1.1 and only summarized below.

From Frisén (1986) it follows that the ML-estimator of μ_s^{Cj}, $j \in \{1,2, ...,s-1\}$ has the property that either

$\hat{\mu}_s^{Cj}(s) = X(s)$

or

$$\hat{\mu}_s^{Cj}(s) = \frac{\sum_{i=b}^{s} X(i)}{s - b_j + 1} = \overline{X}(b_j, s) \text{ and } X(s) > \overline{X}(b_j, s),$$

for some $b_j \in \{j, j+1, ..., s-1\}$ and $j \in \{1, 2, ..., s-1\}$.

$$\hat{\mu}_s^{Cj}(s) = X(s) \qquad \Rightarrow Q_s^{Cj} = Q_{s-1}^{Cj}$$

$$\hat{\mu}_s^{Cj}(s) = \overline{X}(b_j;s),\ X(s) > \overline{X}(b_j,s) \Rightarrow Q_s^{Cj} = \sum_{i=1}^{s-1} Q_s^{Cj}(i) + Q_s^{Cj}(s).$$

The inequality $Q_s^{Cj} > Q_{s-1}^{Cj}$ holds since

$$\sum_{i=1}^{s-1} Q_s^{Cj}(i) > \sum_{i=1}^{s-1} Q_{s-1}^{Cj}(i) \quad \text{and} \quad Q_s^{Cj}(s) > 0.$$

For $\hat{\mu}_s^{Cj}(s) = \overline{X}(b_j;s)$ we have

$$Q_s^{Cj} = \sum_{i=1}^{bj-1} \left(X(i) - \hat{\mu}_s^{Cj}(i)\right)^2 + \sum_{i=bj}^{s-1} \left(X(i) - \overline{X}(b_j;s)\right)^2 + \left(X(s) - \overline{X}(b_j;s)\right)^2.$$

The term $\sum_{i=1}^{bj-1}\left(X(i)-\hat{\mu}_s^{Cj}(i)\right)^2$ is independent of $X(s)$ and the terms $\sum_{i=bj}^{s-1}\left(X(i)-\overline{X}(b_j;s)\right)^2$ and $\left(X(s)-\overline{X}(b_j;s)\right)^2$ are respectively decreasing as $X(s)$ is decreasing.

Thus the quadratic deviations $Q_s^{Cj}, j \in \{1, 2, .., s\text{-}1\}$ increases with $X(s)$ and since $X(s)$ increases stochastically with $\mu(s)$, it follows that $Q_s^{Cj}, j \in \{1, 2, .., s\text{-}1\}$ increases stochastically with $\mu(s)$.■

Lemma 1.3: Q_s^{Cs} is independent of $\mu(s)$

Proof: From Frisén (1986) it follows that $\hat{\mu}_s^{Cs}(s) = X(s)$, i.e. the quadratic deviation Q_s^{Cs} is independent of $X(s)$.■

Appendix B. Lemma (2.1.1) – (2.2), used in Theorem 2

In the Lemmas, the following notations are used:

Q_s^{DB} and Q_s^{CjB} denotes the quadratic deviations for case B

and Q_s^{A} and Q_s^{CjA} denotes the quadratic deviations for case A.

$P^A(t_A \le s)$ denotes $P(t_A \le s \mid \mu = \mu^A)$

and $P\ (t_A \le s)$ denotes $P(t_A \le s \mid \mu = \mu^B)$.

$$P^A\left(MSR(s) > k \,\middle|\, \bigcap_{i\le s-1} MSR(i) < k\right) \text{ denotes}$$
$$P\left(MSR(s) > k \,\middle|\, \bigcap_{i\le s-1} MSR(i) < k, \mu = \mu^A\right)$$

$$\text{and } P^B\left(MSR(s) > k \,\middle|\, \bigcap_{i\le s-1} MSR(i) < k\right) \text{ denotes}$$
$$P\left(MSR(s) > k \,\middle|\, \bigcap_{i\le s-1} MSR(i) < k, \mu = \mu^B\right).$$

Lemma 2.1.1: For two unimodal vectors μ^A and μ^B, as defined in Section 3.1, we have that $E[\hat{\mu}_{s-1}^{DA}(s-1) - \hat{\mu}_{s-1}^{DB}(s-1)] < \mu^A(s) - \mu^B(s)$, $\forall s \geq \tau$.

General description of the consequences of the non-parametric estimation procedure

It follows from the results in Barlow et al. (1972) that the estimation procedure for the vector $\hat{\mu}_{\tau+p-2}^{D}$ at time τ+p-2 can be described as first partitioning the time points j = {1, 2, ..., τ+p-2} into $L(\tau$+p-2) sets, { $b_1^{\tau+p-2}$, ..., $b_{L(\tau+p-2)}^{\tau+p-2}$ }, where

$b_1^{\tau+p-2} = \{1, ..., k_1\}$

$b_2^{\tau+p-2} = \{k_1+1, ..., k_2\}$

...

$b_{L(\tau+p-2)}^{\tau+p-2} = \{k_{L(\tau+p-2)-1}+1, ..., \tau+p-2\}$.

Then, for $\forall j \in b_i^{\tau+p-2}$, it holds that $\hat{\mu}_{\tau+p-2}^{D}(j) = \dfrac{\sum_{u \in b_i^{\tau+p-2}} X(u)}{k_i - k_{i-1}}$,

where the partitioning ensures that

$\hat{\mu}_{\tau+p-2}^{D}(b_1^{\tau+p-2}) < \hat{\mu}_{\tau+p-2}^{D}(b_2^{\tau+p-2}) < ... < \hat{\mu}_{\tau+p-2}^{D}(b_{L(\tau+p-2)}^{\tau+p-2})$.

At time τ+p-1, the time points j = {1, 2, ..., τ+p-1} are partioned into $L(\tau$+p-1) sets. It follows from Barlow et al. (1972) that $\hat{\mu}_{\tau+p-1}^{D}(\tau+p-1)$ is a weighted average of $X(\tau$+p-1) and $(h_{\tau+p-1}+1)$ of the sets from time τ+p-2, where $0 \leq (h_{\tau+p-1}+1) \leq L(\tau$+$p$-2). That is $\hat{\mu}_{\tau+p-1}^{D}(\tau+p-1) =$

$$w_{-1}^{\tau+p-1} \cdot x(\tau+p-1) + w_0^{\tau+p-1} \cdot \hat{\mu}_{\tau+p-2}^{D}(b_{L(\tau+p-1)}^{\tau+p-2}) +$$
$$w_1^{\tau+p-1} \cdot \hat{\mu}_{\tau+p-2}^{D}(b_{L(\tau+p-2)-2}^{\tau+p-2}) + ... + w_{h_\tau}^{\tau+p-1} \cdot \hat{\mu}_{\tau+p-2}^{D}(b_{L(\tau+p-2)-h_{\tau+p-1}}^{\tau+p-2}),$$

where $w_{-1}^{\tau+p-1} + ... + w_{h_{\tau+p-1}}^{\tau+p-1} = 1$.

$E[h_{\tau+p-1}]$ depends on $X(\tau$+p-1), so that

$E[h_{\tau+p-1}]$ increases as $E[X(\tau$+p-1)] decreases.

Since $\hat{\mu}_{\tau+p-2}^{D}(\tau+p-2)$ is independent of $X(\tau$+p-1), it follows that

$E[h_{\tau+p-1}]$ increases as $E[\hat{\mu}_{\tau+p-2}^{D}(\tau+p-2) - \mu(\tau$+$p$-1)] increases.

The estimates $\hat{\mu}^{DA}_{\tau+p-1}(\tau+p-1)$ and $\hat{\mu}^{DB}_{\tau+p-1}(\tau+p-1)$ are expressed as

$$\hat{\mu}^{DA}_{\tau+p-1}(\tau+p-1) =$$

$$w^{A,\tau+p-1}_{-1} \cdot x^{A}(\tau+p-1) + \sum_{i=0}^{h^{A}_{\tau+p-1}} w^{A,\tau+p-1}_{i} \cdot \hat{\mu}^{DA}_{\tau+p-2}(b^{A,\tau+p-2}_{L(\tau+p-2)-i})$$

and $\hat{\mu}^{DB}_{\tau+p-1}(\tau+p-1) =$

$$w^{B,\tau+p-1}_{-1} \cdot x^{B}(\tau+p-1) + \sum_{i=0}^{h^{B}_{\tau+p-1}} w^{B,\tau+p-1}_{i} \cdot \hat{\mu}^{D}_{\tau+p-2}(b^{B,\tau+p-2}_{L(\tau+p-2)-i}),$$

where $\exists$ i: $w^{A,\tau+p-1}_{i} \neq w^{B,\tau+p-1}_{i}$, i = {-1, 0, 1, ...}.

In order to make the comparison easy, $\hat{\mu}^{DA}_{\tau+p-1}(\tau+p-1)$ and $\hat{\mu}^{DB}_{\tau+p-1}(\tau+p-1)$ are expressed using the same weights:

$$\hat{\mu}^{DA}_{\tau+p-1}(\tau+p-1) =$$

$$w^{A,\tau+p-1}_{-1} \cdot x^{A}(\tau+p-1) + \sum_{i=0}^{h^{A}_{\tau+p-1}} w^{A,\tau+p-1}_{i} \cdot \hat{\mu}^{DA}_{\tau+p-2}(b^{A,\tau+p-2}_{L(\tau+p-2)-i})$$

and $\hat{\mu}^{DB}_{\tau+p-1}(\tau+p-1) =$

$$w^{A,\tau+p-1}_{-1} \cdot x^{B}(\tau+p-1) + \sum_{i=0}^{h^{A}_{\tau+p-1}} w^{A,\tau+p-1}_{i} \cdot \hat{\mu}^{DB}_{\tau+p-2}(b^{A,\tau+p-2}_{L(\tau+p-2)-i}) + g_{\tau+p-1},$$

where $g_{\tau+p-1} =$

$$\left(w^{B,\tau+p-1}_{-1} - w^{A,\tau+p-1}_{-1}\right) \cdot x^{B}(\tau+p-1) +$$

$$\sum_{i=0}^{h^{B}_{\tau+p-1}} \left(w^{B,\tau+p-1}_{i} - w^{A,\tau+p-1}_{i}\right) \cdot \hat{\mu}^{DB}_{\tau+p-2}(b^{B,\tau+p-2}_{L(\tau+p-2)-i}).$$

The expected difference between the two estimates equals

$$\mathrm{E}[\,\hat{\mu}^{DA}_{\tau+p-1}(\tau+p-1) - \hat{\mu}^{DB}_{\tau+p-1}(\tau+p-1)\,] =$$

$$w^{A,\tau+p-1}_{-1} \cdot \mathrm{E}[\,X^{A}(\tau+p-1) - X^{B}(\tau+p-1)\,] +$$

$$\sum_{i=0}^{h^{A}_{\tau+p-1}} w^{A,\tau+p-1}_{i} \cdot \mathrm{E}[\,\hat{\mu}^{DA}_{\tau+p-2}(b^{\tau+p-2}_{L(\tau+p-2)-i}) - \hat{\mu}^{DB}_{\tau+p-2}(b^{\tau+p-2}_{L(\tau+p-2)-i})\,] - g_{\tau+p-1} \cdot$$

Look at the sum

$$\sum_{i=0}^{h^A_{\tau+p-1}} w_i^{A,\tau+p-1} \cdot \mathrm{E}[\,\hat{\mu}^{DA}_{\tau+p-2}(b^{\tau+p-2}_{L(\tau+p-2)-i}) - \hat{\mu}^{DB}_{\tau+p-2}(b^{\tau+p-2}_{L(\tau+p-2)-i})\,].$$

The estimation procedure is such that for every $\hat{\mu}^{D}_{\tau+p-2}(b_j^{\tau+p-2})$, only observations from time point 1 up to the last time point of the set $b_j^{\tau+p-2}$ can be included in $\hat{\mu}^{D}_{\tau+p-2}(b_j^{\tau+p-2})$.

Since

$(\mu^A(\tau+p-1) - \mu^B(\tau+p-1))$ increases as p increases, $p \geq 1$

it follows that the largest possible expected difference in the sum is

$$\mathrm{E}[\,\hat{\mu}^{DA}_{\tau+p-2}(b^{\tau+p-2}_{L(\tau+p-2)}) - \hat{\mu}^{DB}_{\tau+p-2}(b^{\tau+p-2}_{L(\tau+p-2)})\,].$$

Therefore the following inequality holds for the expected difference:

$$\mathrm{E}[\,\hat{\mu}^{DA}_{\tau+p-1}(\tau+p-1) - \hat{\mu}^{DB}_{\tau+p-1}(\tau+p-1)\,] \leq$$

$$w_{-1}^{A,\tau+p-1} \cdot \mathrm{E}[\,X_A(\tau+p-1) - X_B(\tau+p-1)\,] +$$

$$\mathrm{E}[\,\hat{\mu}^{DA}_{\tau+p-2}(b^{\tau+p-2}_{L(\tau+p-2)}) - \hat{\mu}^{DB}_{\tau+p-2}(b^{\tau+p-2}_{L(\tau+p-2)})\,] \cdot \sum_{i=0}^{h^A_{\tau+p-1}} w_i^{A,\tau+p-1} - g_{\tau+p-1}$$

i.e.

$$\mathrm{E}[\,\hat{\mu}^{DA}_{\tau+p-1}(\tau+p-1) - \hat{\mu}^{DB}_{\tau+p-1}(\tau+p-1)\,] \leq$$

$$w_{-1}^{A,\tau+p-1} \cdot \left(\mu_A(\tau+p-1) - \mu_B(\tau+p-1)\right) +$$
$$\mathrm{E}[\,\hat{\mu}^{DA}_{\tau+p-2}(b^{\tau+p-2}_{L(\tau+p-2)}) - \hat{\mu}^{DB}_{\tau+p-2}(b^{\tau+p-2}_{L(\tau+p-2)})\,] \cdot \left(1 - w_{-1}^{A,\tau+p-1}\right) - g_{\tau+p-1}.$$

$g_{\tau+p-1}$ depends on the relation between $h^A_{\tau+p-1}$ and $h^B_{\tau+p-1}$.

For $h^A_{\tau+p-1} = h^B_{\tau+p-1}$, then $w_i^{B,\tau+p-1} = w_i^{A,\tau+p-1}$, i = {-1, 0, 1, ...} and therefore

$$w_{-1}^{A,\tau+p-1} \cdot x_B(\tau+p-1) + \sum_{i=0}^{h^A_{\tau+p-1}} w_i^{A,\tau+p-1} \cdot \hat{\mu}^{DB}_{\tau+p-2}(b^{A,\tau+p-2}_{L(\tau+p-2)-i}) =$$

$$w_{-1}^{B,\tau+p-1} \cdot x^B(\tau+p-1) + \sum_{i=0}^{h^B_{\tau+p-1}} w_i^{B,\tau+p-1} \cdot \hat{\mu}^{DB}_{\tau+p-2}(b^{A,\tau+p-2}_{L(\tau+p-2)-i})$$

and then

$$g_{\tau+p-1} = 0.$$

For $h^A_{\tau+p-1} < h^B_{\tau+p-1}$, then $\exists$ i: $w_i^{A,\tau+p-1} \neq w_i^{B,\tau+p-1}$, i = {-1, 0, 1, ...} and therefore

$$w_{-1}^{B,\tau+p-1} \cdot x^B(\tau+p-1) + \sum_{i=0}^{h^B_{\tau+p-1}} w_i^{B,\tau+p-1} \cdot \hat{\mu}^{DB}_{\tau+p-2}(b^{B,\tau+p-2}_{L(\tau+p-2)-i}) >$$

$$w_{-1}^{,\tau+p-1} \cdot x^B(\tau+p-1) + \sum_{i=0}^{h^A_{\tau+p-1}} w_i^{,\tau+p-1} \cdot \hat{\mu}^{DB}_{\tau+p-2}(b^{,\tau+p-2}_{L(\tau+p-2)-i})$$

and then

$$g_{\tau+p-1} > 0.$$

Proof Lemma 2.1.1:

Here starts the proof of Lemma 2.1.1. We use induction to prove that
$E[\hat{\mu}^{DA}_{s-1}(s-1) - \hat{\mu}^{DB}_{s-1}(s-1)] < \mu^A(s) - \mu^B(s)$, for $s \geq \tau$.

The estimation procedure is such that
$E[h_{\tau+p-1}]$ increases as $(\hat{\mu}^D_{\tau+p-2}(\tau+p-2) - \mu(\tau+p-1))$ increases.

Induction: if Lemma 2.1.1 is valid at time $s=\tau+p-1$, then the following holds

$$E[\hat{\mu}^{DA}_{\tau+p-2}(\tau+p-2) - \hat{\mu}^{DB}_{\tau+p-2}(\tau+p-2)] < \mu^A(\tau+p-1) - \mu^B(\tau+p-1),$$

and then

$$E[h^A_{\tau+p-1}] < E[h^B_{\tau+p-1}]$$

since

$$E[\hat{\mu}^{DA}_{\tau+p-2}(\tau+p-2) - X_A(\tau+p-1)] <$$
$$E[\hat{\mu}^{DB}_{\tau+p-2}(\tau+p-2) - X^B(\tau+p-1)].$$

The largest possible expected difference is

$$E[\hat{\mu}^{DA}_{\tau+p-1}(\tau+p-1) - \hat{\mu}^{DB}_{\tau+p-1}(\tau+p-1)] \leq$$

$$w^{A,\tau+p-1}_{-1} \cdot (\mu_A(\tau+p-1) - \mu_B(\tau+p-1)) +$$
$$E[\hat{\mu}^{DA}_{\tau+p-2}(b^{\tau+p-2}_{L(\tau+p-2)}) - \hat{\mu}^{DB}_{\tau+p-2}(b^{\tau+p-2}_{L(\tau+p-2)})] \cdot \left(1 - w^{A,\tau+p-1}_{-1}\right) - g_{\tau+p-1}.$$

i.e. $E[\hat{\mu}^{DA}_{\tau+p-1}(\tau+p-1) - \hat{\mu}^{DB}_{\tau+p-1}(\tau+p-1)] \leq$

$$w^{A,\tau+p-1}_{-1} \cdot \left(\mu^A(\tau+p-1) - \mu^B(\tau+p-1)\right) +$$
$$E[\hat{\mu}^{DA}_{\tau+p-2}(\tau+p-2) - \hat{\mu}^{DB}_{\tau+p-2}(\tau+p-2)] \cdot \left(1 - w^{A,\tau+p-1}_{-1}\right) - g_{\tau+p-1}.$$

Again, if Lemma 2.1.1 is valid for s=τ+p-1, then the following inequality holds

$E[\hat{\mu}^{DA}_{\tau+p-2}(\tau+p-2) - \hat{\mu}^{DB}_{\tau+p-2}(\tau+p-2)] < \mu^A(\tau+p-1) - \mu^B(\tau+p-1)$

and then the following inequality holds for the expected difference:

$E[\hat{\mu}^{DA}_{\tau+p-1}(\tau+p-1) - \hat{\mu}^{DB}_{\tau+p-1}(\tau+p-1)] \leq$

$$w^{A,\tau+p-1}_{-1} \cdot \left(\mu^A(\tau+p-1) - \mu^B(\tau+p-1)\right) +$$
$$\left(\mu^A(\tau+p-1) - \mu^B(\tau+p-1)\right) \cdot \left(1 - w^{A,\tau+p-1}_{-1}\right) - g_{\tau+p-1},$$

where $g_{\tau+p-1} > 0$, if Lemma 2.1.1 is valid at time $s=\tau+p-1$.

For $g_{\tau+p-1} > 0$, the following inequality holds for the expected difference:

$E[\hat{\mu}^{DA}_{\tau+p-1}(\tau+p-1) - \hat{\mu}^{DB}_{\tau+p-1}(\tau+p-1)] \leq$

$$w^{A,\tau+p-1}_{-1} \cdot \left(\mu^A(\tau+p-1) - \mu^B(\tau+p-1)\right) +$$
$$\left(\mu^A(\tau+p-1) - \mu^B(\tau+p-1)\right) \cdot \left(1 - w^{A,\tau+p-1}_{-1}\right).$$

Since $w^{A,\tau+p-1}_{-1} \leq 1$, the following inequality holds for the expected difference:

$E[\hat{\mu}^{DA}_{\tau+p-1}(\tau+p-1) - \hat{\mu}^{DB}_{\tau+p}(\tau+p-1)] \leq \mu^A(\tau+p-1) - \mu^B(\tau+p-1)$.

Since $(\mu^A(\tau+p-1) - \mu^B(\tau+p-1)) < (\mu^A(\tau+p) - \mu^B(\tau+p))$, the following inequality holds

$$E[\hat{\mu}^{DA}_{\tau+p-1}(\tau+p-1) - \hat{\mu}^{DB}_{\tau+p-1}(\tau+p-1)] < \mu^A(\tau+p) - \mu^B(\tau+p).$$

Thus, Lemma 2.1.1 holds for $s=\tau+p$ if it holds for $s=\tau+p-1$.
Lemma 2.1.1 holds for $s=\tau$ by Lemma 1.1, thus it holds for $\{s=\tau+1, s=\tau+2, ...\}$, all $s \geq \tau$. ■

Lemma 2.1.2: For two unimodal vectors μ^A and μ^B, as defined in Section 3.1, we have that (Q_s^{DB} - Q_{s-1}^{DB}) is stochastically larger than (Q_s^{DA} - Q_{s-1}^{DA}), for $s \geq \tau$.

Proof Lemma 2.1.2:

As written in Lemma 1.1, the ML-estimator has the property that

when $X(s) > \hat{\mu}_{s-1}^D(s-1)$, then $Q_s^D = Q_{s-1}^D$

and when $X(s) < \hat{\mu}_{s-1}^D(s-1)$, then $Q_s^D > Q_{s-1}^D$, $\forall\, s$.

This could be expressed as

when $\hat{\mu}_{s-1}^D(s-1) - X(s) > 0$, then $Q_s^D > Q_{s-1}^D$.

From Lemma 1.1 it follows that

Q_s^D increases as $X(s)$ decreases,

i.e. $Q_{s-1}^D + (Q_s^D - Q_{s-1}^D)$ increases as $X(s)$ decreases,

i.e. $Q_{s-1}^D + (Q_s^D - Q_{s-1}^D)$ increases as $(\hat{\mu}_{s-1}^D(s-1) - X(s))$ increases, $\forall\, s$.

Since Q_{s-1}^D and $\hat{\mu}_{s-1}^D(s-1)$ are independent of $X(s)$, it follows that

$X(s)$ decreases $\Rightarrow$

$(\hat{\mu}_{s-1}^D(s-1) - X(s))$ increases $\Rightarrow$

$(Q_s^D - Q_{s-1}^D)$ increases.

The following inequality was proven in Lemma 2.1.1:

$E[\hat{\mu}_{s-1}^{DA}(s-1) - \hat{\mu}_{s-1}^{DB}(s-1)] < \mu^A(s) - \mu^B(s)$,

i.e. $E[\hat{\mu}_{-1}^A(s-1) - X^A(s)] < E[\hat{\mu}_{-1}(s-1) - X^B(s)]$.

It follows from Lemma 1.1 that

$X(s)$ decreases $\Rightarrow$

$(\hat{\mu}_{-1}(s-1) - X(s))$ increases $\Rightarrow$

$(Q - Q_{-1})$ increases.

It follows from Lemma 1.1 that $(Q - Q_{-1})$ increases stochastically with $E[\hat{\mu}_{-1}(s-1) - X(s)]$.

Since $E[\hat{\mu}_{-1}^A(s-1) - X^A(s)] < E[\hat{\mu}_{-1}^B(s-1) - X^B(s)]$, it follows that

$(Q^B - Q_{-1}^B)$ is stochastically larger than $(Q^A - Q_{-1}^A)$. ■

Lemma 2.2: For two unimodal vectors μ^A and μ^B, as defined in Section 3.1, we have that (Q_s^{CjB} - Q_{s-1}^{CjB}) is stochastically smaller than (Q_s^{CjA} - Q_{s-1}^{CjA}), $j=\{1, 2, ..., s\text{-}1\}$, for $s \geq \tau$.

Proof Lemma 2.2:
The proof is made correspondingly to the proof of Lemma 2.1.1 and 2.1.2 and therefore is not given here.

Appendix C. Proof of Theorem 2.

Theorem 2: For two unimodal vectors μ^A and μ^B, as defined above, we have that $P^B(t_A \leq s) > P^A(t_A \leq s)$, $\forall\ s \geq \tau$.

General description of the alarm statistic, MSR(s).

The alarm statistic at time s is a sum of s components

$$\mathrm{MSR}(s) = \exp\left(\frac{Q_s^D - Q_s^{C1}}{2}\right) + \ldots + \exp\left(\frac{Q_s^D - Q_s^{Cs}}{2}\right)$$

First we give a general description of the dependence between MSR(s-1) and MSR(s).

MSR(s-1) is a sum of (s-1) components, denoted MSR_1(s-1), ..., MSR_{s-1}(s-1), so that

$$\mathrm{MSR}(s\text{-}1) = \mathrm{MSR}_1(\text{s-}1) + \ldots + \mathrm{MSR}_{s-1}(\text{s-}1)$$

and MSR(s) is a weighted sum of the (s-1) components of MSR(s-1), plus one additional component:

$$\mathrm{MSR}(s) =$$

$$\exp\left(\frac{(Q_s^D - Q_{s-1}^D) - (Q_s^{C1} - Q_{s-1}^{C1})}{2}\right) MSR_1(s-1) + \ldots +$$

$$\exp\left(\frac{(Q_s^D - Q_{s-1}^D) - (Q_s^{Cs-1} - Q_{s-1}^{Cs-1})}{2}\right) MSR_{s-1}(s-1) +$$

$$\exp\left(\frac{Q_s^D - Q_s^{Cs}}{2}\right).$$

Denote

$$\frac{(Q_s^D - Q_{s-1}^D) - (Q_s^{Cj} - Q_{s-1}^{Cj})}{2} \text{ by } \phi_s^j$$

and $\frac{Q_s^D - Q_s^{Cs}}{2}$ by δ_s.

By the notation above MSR(s) is expressed as

$$\text{MSR}(s) = \exp\left(\phi_s^1\right) \cdot MSR_1(s-1) + \ldots + \exp\left(\phi_s^{s-1}\right) \cdot MSR_{s-1}(s-1) + \exp(\delta_s).$$

Analogically, MSR(τ+p), p={0, 1, 2, ...}, can be expressed as a weighted sum of the components of MSR(τ-1) plus p additional components.

MSR(τ+p) =

$$\exp\left(\phi_\tau^1 + \ldots + \phi_{\tau+p}^1\right) \cdot MSR^1(\tau-1) + \ldots + \exp\left(\phi_\tau^{\tau-1} + \ldots + \phi_{\tau+p}^{\tau-1}\right) \cdot MSR^{\tau-1}(\tau-1) +$$
$$\exp\left(\phi_{\tau+1}^{\tau} + \ldots + \phi_{\tau+p}^{\tau}\right) \cdot \exp(\delta_\tau) + \exp\left(\phi_{\tau+2}^{\tau+1} + \ldots + \phi_{\tau+p}^{\tau+1}\right) \cdot \exp(\delta_{\tau+1}) + \ldots +$$
$$\exp\left(\phi_{\tau+p}^{\tau+p-1}\right) \cdot \exp\left(\delta_{\tau+p-1}\right) + \exp\left(\delta_{\tau+p}\right).$$

From Lemma 2.1.2 and 2.2 we have that, for $s \geq \tau$,

$(Q_s^{DB} - Q_{s-1}^{DB})$ is stochastically larger than $(Q_s^{DA} - Q_{s-1}^{DA})$
and $(Q_s^{CjB} - Q_{s-1}^{CjB})$ is stochastically smaller than $(Q_s^{CjA} - Q_{s-1}^{CjA}), j = \{1, \ldots, s\text{-}1\}$.

From this it follows that, for $s \geq \tau$,
$(Q_s - Q_{s-1}) - (Q_s^{Cj} - Q_{s-1}^{Cj})$ is stochastically larger than
$(Q_s^A - Q_{s-1}^A) - (Q_s^{CjA} - Q_{s-1}^{CjA}), j = \{1, \ldots, s\text{-}1\}$

i.e.

ϕ_s^j is stochastically larger than $\phi_s^{jA}, j = \{1, \ldots, s\text{-}1\}, s \geq \tau$.

The estimation procedure is such that $Q_s^{Cs} = Q_{s-1}^D$. Therefore

$\delta_s = Q_s^D - Q_s^{Cs} = Q_s^D - Q_{s-1}^D$.

By Lemma 2.1.2, we have that, for $s \geq \tau$,

$(Q_s^D - Q_{s-1}^D)$ is stochastically larger than $(Q_s^D - Q_{s-1}^D)$

i.e.

δ_s^B is stochastically larger than $\delta_s^A, s \geq \tau$.

Proof Theorem 2:

Induction: if Theorem 2 is valid for $s= \tau+p-1$, then the following inequality holds

$$P^B(t_A \le \tau + p - 1) > P^A(t_A \le \tau + p - 1)$$

that is

$$P^A(t_A > \tau + p - 1) > P^B(t_A > \tau + p - 1).$$

This is equivalent to the event that the alarm statistic, at no time point, has exceeded the alarm limit, i.e.

$$P^A\left(\bigcap_{i\le\tau+p-1} MSR(i) < k\right) > P^B\left(\bigcap_{i\le\tau+p-1} MSR(i) < k\right).$$

We will start from the following conditional probability

$$P\left(MSR(\tau + p) < k \middle| \bigcap_{i\le\tau+p-1} MSR(i) < k\right).$$

Denote

$\left\{\bigcap_{i\in(1:j)} MSR(j) < k\right\}$ by $M_{1:j}$.

The conditional probability

$$P\left(MSR(\tau + p) < k \middle| M_{1:\tau+p-1}\right)$$

is expressed as

$$\frac{P\left(\left\{\bigcap_{i\in\{0:p\}} MSR(\tau + i) < k\right\} \middle| \{M_{1:\tau-1}\}\right) \cdot P(M_{1:\tau-1})}{P(M_{1:\tau+p-1})}.$$

Look at the event $\left\{\bigcap_{i\in\{0:p\}} MSR(\tau + i) < k\right\}$.

The alarm statistic MSR($\tau+i$), $i = \{0, 1, ..., p\}$ is expressed as

MSR($\tau+i$) =

$$\exp\left(\phi_{\tau}^{1}+...+\phi_{\tau+i}^{1}\right)\cdot MSR^{1}(\tau-1)+...+\exp\left(\phi_{\tau}^{\tau-1}+...+\phi_{\tau+i}^{\tau-1}\right)\cdot MSR^{\tau-1}(\tau-1)+$$
$$\exp\left(\phi_{\tau+1}^{\tau}+...+\phi_{\tau+1}^{\tau}\right)\cdot\exp(\delta_{\tau})+\exp\left(\phi_{\tau+2}^{\tau+1}+...+\phi_{\tau+p}^{\tau+1}\right)\cdot\exp(\delta_{\tau+1})+...+$$
$$\exp\left(\phi_{\tau+i}^{\tau+i-1}\right)\cdot\exp(\delta_{\tau+i-1})+\exp(\delta_{\tau+i}).$$

From the results in Lemma 2.1.2 and 2.2 it follows that

ϕ_s^{jA} is stochastically smaller than ϕ_s^{jB}, $j = \{1,..., s\text{-}1\}$, $s\geq\tau$,
and
δ_s^A is stochastically smaller than δ_s^B, $s\geq\tau$.

Therefore it follows that the alarm statistic MSR($\tau+i$), at every time $i=\{0, 1, ..., p\}$, is stochastically smaller for case A than for case B, i.e.

$$P^{A}(MSR(\tau+i)<k)>P^{B}(MSR(\tau+i)<k),\ i=\{0,1,...,p\}.$$

From this it follows that the event $\left\{\bigcap_{i\in\{0:p\}}MSR(\tau+i)<k\right\}$ is stochastically smaller for case A than for case B, i.e.

$$P^{A}\left(\left\{\bigcap_{i\in\{0:p\}}MSR(\tau+i)<k\right\}\right)>P^{B}\left(\left\{\bigcap_{i\in\{0:p\}}MSR(\tau+i)<k\right\}\right).$$

Since $P^{B}(M_{1:\tau-1})=P^{A}(M_{1:\tau-1})$ we have

$$P^{A}\left(\left\{\bigcap_{i\in\{0:p\}}MSR(\tau+i)<k\right\}\middle|\{M_{1:\tau-1}\}\right)>$$
$$P^{B}\left(\left\{\bigcap_{i\in\{0:p\}}MSR(\tau+i)<k\right\}\middle|\{M_{1:\tau-1}\}\right)$$

and further that

$$P\left(\left\{\bigcap_{i\in\{0:p\}}MSR(\tau+i)<k\right\}\cap\{M_{1:\tau-1}\}\right)>$$
$$P^{B}\left(\left\{\bigcap_{i\in\{0:p\}}MSR(\tau+i)<k\right\}\cap\{M_{1:\tau-1}\}\right).$$

The inequality above is equivalent to

$$P^{A}\left(\bigcap_{i\leq\tau+p}MSR(i)<k\right)>P^{B}\left(\bigcap_{i\leq\tau+p}MSR(i)<k\right)$$

i.e.

$$P^A(t_A > \tau + p) > P^B(t_A > \tau + p)$$

i.e.

$$P^B(t_A \leq \tau + p) > P^A(t_A \leq \tau + p).$$

Thus Theorem 2 holds for $s=\tau+p$ if it holds for $s=\tau+p-1$.

It holds for $s=\tau$ by Corollary 1, thus it holds for $\{s=\tau+1, s=\tau+2, ...\}$, all $s \geq \tau$. ■

Research Report

2000:1	Hatemi-J, A. & Shukur, G.:	Multivariate based causality tests of twin deficits in the US.
2000:2	Petzold, M.:	Aspects on tolerance limit estimation - some common approaches and flexible modeling.
2000:3	Andersson, L.:	Statistical test of the existence of a turning point.
2000:4	Almasri, A.:	Some aspects of wavelet analysis in time series.
2000:5	Järpe, E.:	Surveillance of spatio-temporal patterns. Change of interaction in an Ising dynamic model.
2000:6	Järpe, E.:	Detection of environmental catastrophes.
2001:1	Holgersson, H.E.T.:	On assessing multivariate normality.
2001:2	Sonesson, C. & Bock, D.:	Statistical issues in public health monitoring – A review and discussion.