



Research Report
Statistical Research Unit
Department of Economics
Göteborg University
Sweden

Evaluations of likelihood based surveillance of volatility

David Bock

Research Report 2007:9
ISSN 0349-8034

Mailing address:	Fax	Phone	Home Page:
Statistical Research Unit P.O. Box 640 SE 405 30 Göteborg Sweden	Nat: 031-786 12 74	Nat: 031-786 00 00 Int: +46 31 786 12 74	http://www.statistics.gu.se/

EVALUATIONS OF LIKELIHOOD BASED SURVEILLANCE OF VOLATILITY

By David Bock

Statistical Research Unit, Göteborg University

ABSTRACT

The volatility of asset returns are important in finance. Different likelihood based methods of statistical surveillance for detecting a change in the variance are evaluated. The differences are how the partial likelihood ratios are weighted. The full likelihood ratio, Shiryaev-Roberts, Shewhart and the CUSUM methods are derived in case of an independent and identically distributed Gaussian process. The behavior of the methods is studied both when there is no change and when the change occurs at different time points. The false alarms are controlled by the median run length. Differences and limiting equalities of the methods are shown. The performances when the process parameters for which the methods are optimized for differ from the true values of the parameters are evaluated. The methods are illustrated on a period of Standard and Poor's 500 stock market index.

Key Words: surveillance; statistical process control; monitoring; likelihood ratio; Shewhart; CUSUM.

1	INTRODUCTION	1
2	THE CHANGE-POINT PROBLEM	3
3	MEASURES OF EVALUATION AND OPTIMALITY CRITERIA	4
4	METHODS FOR SURVEILLANCE.....	5
	4.1 Suggested statistics under surveillance.....	5
	4.2 Likelihood based methods	5
	4.3 Limiting equalities	7
5	A MONTE CARLO STUDY	8
	5.1 In-control properties.....	9
	5.2 Out-of-control properties	10
	5.3 The trust of alarms	11
6	ILLUSTRATIVE EXAMPLE.....	13
7	CONCLUDING REMARKS	14
	REFERENCES	15

1 INTRODUCTION

Timely detection of an important change in stochastic processes is important in many areas. In finance, detecting changes in assets prices or returns are important for investment decisions and Shiryaev (2002) demonstrated that change-points might induce arbitrage opportunities. The surveillance of business cycles, treated in the special issue (no. 3/4 1993) of Journal of Forecasting, Andersson et al. (2004),

Andersson et al. (2005) and Andersson et al. (2006) is another important application. In quality control (Wetherhill and Brown (1991)), we may aim at timely detecting contaminated products in a manufacturing process. In medicine it is important to detect e.g. intrauterine growth retardation (Petzold et al. (2003)) or an increased incidence of a disease (Sonesson and Bock (2003)). Different medical applications are described in the special issue (no. 3 1989) of *Statistics in Medicine*.

Since timeliness is important, the data should not be analyzed in a retrospective setting but in a prospective framework where data are analyzed online and sequential decisions are made. In the inference situation of surveillance repeated decisions are made, the sample size is increasing and the null hypothesis is never accepted. For general reviews on statistical surveillance, see Frisén and de Maré (1991), Yashchin (1993), Srivastava and Wu (1993), Lai (1995), Frisén and Wessman (1999) and Frisén (2003).

Many methods for surveillance are in one way or another based on likelihood ratios. Likelihood ratio based methods are known to possess several optimality properties and the different methods are suitable for different situations. The methods has mostly been constructed and evaluated in a situation where the aim is to detect a change in the level of the process. Increasing attention has however been given to the monitoring of the variance (or the standard deviation).

Detecting changes in the volatility of asset returns are important in e.g. portfolio management, see Severin and Schmid (1998), Severin and Schmid (1999), Schipper and Schmid (2001a) and Schipper and Schmid (2001b) where several surveillance methods were compared with respect to detecting changes in GARCH (generalized autoregressive conditional heteroscedasticity) processes, which are used to describe volatility in financial markets.

The aim of this paper is to construct and evaluate likelihood based methods for detecting a change in volatility. Assuming a GARCH process might be reasonable in a financial setting but since no explicit expression for the univariate marginal distribution of a GARCH process is known (Schipper and Schmid (2001b)) constructing the required likelihood is not possible. Therefore an independent Gaussian process is studied here instead, as in most of the literature. The methods studied are the full likelihood ratio (LR), Shiryaev-Roberts (SR), Shewhart and the CUSUM methods, presented in section 4. The methods differ in what way the different partial likelihood ratios are weighted and they depend on different number of process parameters.

These methods were studied in Frisén and Wessman (1999) and Järpe and Wessman (2000) for the same process as here but a change in the level. In case of a change in the variance, earlier studies have been made of the Shewhart (see e.g. Reynolds and Soumbos (2001)), CUSUM (e.g. Srivastava (1997) and Acosta-Mejia et al. (1999)) and SR (Srivastava and Chow (1992)) methods.

Different variants of the EWMA (exponentially weighted moving averages) method suggested by Roberts (1959) are often suggested, see e.g. Crowder and Hamilton (1992), MacGregor and Harris (1993), Acosta-Mejía and Pignatiello (2000) and Schipper and Schmid (2001b). The EWMA method is not likelihood based and is not studied here.

The performance has often been assessed by the average run length when a change happens either immediately or never. The sole use of these two measures has however been criticized and a single measure of performance is not always enough but evaluations of different properties might be necessary, as pointed out by several authors, e.g. Frisén (1992) and Frisén (2003). In Frisén and Wessman (1999) and

Järpe and Wessman (2000) the methods were made comparable by having the same average run length when there is no change. Here the median run length is used.

The different parameters can be chosen to make the methods optimal for specific situations. Since information on the parameters is rarely known in practice, there is a risk of mis-specification. The effect of mis-specifications on the performance of the methods is studied.

As an illustrative example we monitor a period of Standard and Poor's 500 stock market index to investigate whether our procedures could have detected a documented change in volatility.

The plan of this paper is as follows. Notations and specifications are given in section 2. Optimality and measures of evaluation are described in section 3. Methods are described in section 4. Results from a simulation study are given in section 5 and in section 6 the methods are applied in a case study. Concluding remarks are given in section 7.

2 THE CHANGE-POINT PROBLEM

The process under surveillance, denoted by X is, as in most literature on quality control, measured at discrete time points, $t=1, 2, \dots$ and assumed to be independent Gaussian. Both the situation with subgroups that is samples of more than one observation are made at each time and without subgroups that is a single observation is made at each time have been treated in the literature. Often the both location and dispersion are monitored simultaneously.

Here we have a single observation at each time and at an unknown time point, denoted by τ , there is an increase in the variance;

$$V[X(t)] = \begin{cases} \sigma^2, & t < \tau \\ \Delta \cdot \sigma^2, & t \geq \tau \end{cases}$$

where $\Delta > 1$ is the unknown size of the shift. At time $t < \tau$ and $t \geq \tau$ the process is said to be in-control and out-of-control, respectively. The aim is to detect the change as soon as possible after it has occurred. Only one-sided procedures are considered. In quality control the expected value μ and σ^2 are often regarded as unknown and the change point time τ is an unknown non-random parameter. Here σ^2 and μ are considered as known and the change point time τ is a discrete-valued random variable with intensity parameter

$$v_t = P(\tau=t | \tau \geq t).$$

We treat the case of a constant unknown intensity v that is τ has a geometric distribution with density $P(\tau=t) = v \cdot (1-v)^{t-1}$ on $t=1, 2, \dots$ as in e.g. Shiryaev (1963) and Frisén and Wessman (1999). Without loss of generality we take $\mu=0$. In those methods where it is required, the unknown parameters Δ and v are replaced by values d and v , respectively. The values are chosen to be relevant for the problem at hand and the methods are optimized for these values.

At each decision time $s=1, 2, \dots$, we want to discriminate between $C(s)$ and $D(s)$ where $C(s)$ is the critical event implying that the process is out-of-control and $D(s)$

implies that the process is in-control. The $C(s)$ and $D(s)$ can be specified in various ways and different methods are optimal for different specifications. Sometimes it is important to see whether there has been a change since the start of the surveillance and then $C(s)=\{\tau \leq s\}$ and $D(s)=\{\tau > s\}$.

An alarm set $A(s)$ is constructed, with the property that as soon as X_s belongs to $A(s)$ we infer that $C(s)$ has occurred. Usually the alarm set consists of an alarm statistic $p(x_s)$ and an alarm limit $g(s)$, where the time of an alarm, t_A , is defined as

$$t_A = \min\{s: p(x_s) > g(s)\}.$$

3 MEASURES OF EVALUATION AND OPTIMALITY CRITERIA

A desirable property of a method is that it detects a change quickly without having too many false alarms. We must however face a trade off between false alarms and the ability to detect a change. Likewise traditional hypothesis testing optimality of surveillance methods is assessed by the detection ability given a controlled error rate. However, as opposed to hypothesis testing is surveillance characterized by repeated decisions. Consequently, measures such as the significance level and the power need to be generalized to consider the sequential aspect.

Chu et al. (1996) advocated a controlled the probability of any false alarm during an infinitely long surveillance period, $\lim_{i \rightarrow \infty} P(t_A \leq i | D) < 1$. This is convenient since ordinary statements of hypothesis testing can be made. It was however pointed out by Pollak and Siegmund (1975) and Frisén (1994) that the ability to detect a change deteriorates rapidly with the time of the change. Consequences of this were illustrated in Bock (2006).

A commonly used measure to summarize and control the false alarms is by the average run length, $ARL^0 = E[t_A | D]$. Hawkins (1992) and Gan (1993) suggest that the control is made by the median run length, $MRL^0 = \text{Median}[t_A | D]$ as it has easier interpretations for skewed distributions and much shorter computer time for calculations. A third measure is the probability of a false alarm, $PFA = P(t_A < \tau) = E_\tau[P(t_A < \tau | \tau = t)]$, which can be thought of as a characteristic for surveillance corresponding to the level of significance for hypothesis testing (Järpe and Wessman (2000)).

The timeliness of motivated alarms can be reflected by the average run length given an immediate change, $ARL^1 = E[t_A | \tau = 1]$. This is the most commonly used measure but it is relevant to consider other change point times as well, as will be discussed later. The ability to detect a change within m time units from τ is reflected by the probability of successful detection, $PSD(m, t) = P(t_A - \tau \leq m | t_A \geq \tau, \tau = t)$, $m=0, 1, \dots$. It was suggested by Frisén (1992) and is an important measure if there is limited time available for rescuing action, e.g. in the surveillance of the fetal heart rate during labor or intrauterine growth retardation. Another measure is the conditional expected delay $CED(t) = E[t_A - \tau | t_A \geq \tau, \tau = t]$. The delay is summarizing with respect to the distribution of τ by $ED = E_\tau[ED(\tau)]$ where $ED(t) = CED(t) \cdot P(t_A \geq \tau)$. An important aspect when evaluating a method is the trust you should have in an alarm at a specific time. The predictive value of an alarm at time t , $PV(t) = P(\tau \leq t | t_A = t)$, suggested by Frisén (1992) reflects the trust of an alarm.

The most commonly used optimality criteria is minimal ARL^1 for a fixed ARL^0 . In the literature on control charts for the variance, this criterion has been used with only one exception (Hawkins and Zamba (2005)). This criterion might be suitable in an industrial manufacturing process where one considers various start-up problems. An advantage is that the criterion does not require an assumption regarding the distribution of τ but Frisén (2003) and Frisén and Sonesson (2003) has questioned it as a formal criterion.

In the utility function suggested by Shiryaev (1963) the gain of an alarm is a linear function of the expected delay. The loss of a false alarm is an arbitrary function of the same difference. The criterion of maximization of the expected utility, where the expectation is taken with respect to τ , is often referred to as the ED criterion (see e.g. Frisén (2003)), since the expected delay is to be minimized. Bock et al. (2006) demonstrated that for certain assumptions regarding the price on assets, fulfilling of the ED criterion is equivalent to maximizing the expected return.

When the worst possible case is important, the minimax criteria of Moustakides (1986) can be used. The criterion is minimal CED given the worst possible value of τ and the worst possible outcome of $X_{\tau-1}$, given a fixed ARL^0 . As only the worst possible value of CED is used, a distribution of τ is not required.

4 METHODS FOR SURVEILLANCE

4.1 *Suggested statistics under surveillance*

For the situation specified in section 2 $(X(t)-\mu)^2$, $t=1, 2, \dots, s$ are sufficient for the problem as will be seen in the next section. Often a transformation of the estimated variance at each time is used in the alarm statistic. Different transformations have different motivations. Often the transformation is made such that the variable under surveillance is (approximately) Gaussian and standard charts for Gaussian variables therefore can be used. Examples such of transformations are the logarithm of the subgroup standard deviation (Crowder and Hamilton (1992)) and $|X(t)/\sigma^2|^{1/2}$ (Hawkins (1981)).

In the presence of a nuisance parameter using a pivot statistic is often advocated. The sub group range or a moving range or consecutive differences when there are no subgroups have been suggested when μ is unknown as these statistics are robust to changes in μ , see e.g. Page (1963), Rigdon et al. (1994), Acosta-Mejia (1998) and Acosta-Mejia and Pignatiello (2000). For, on the other hand, simultaneous surveillance of μ and the variance by a single statistic, see Domangue and Patch (1991), Chen et al. (2004) and Costa and Rahim (2004).

In Ncube and Li (1999) the values of the EWMA statistic are discretized by a score that is assigned different values depending on the process is within different intervals. The alarm statistic is formed by the cumulative score. This could be motivated from a robustness perspective but implies a suboptimal procedure as a direct loss of information owing to the discretization of the data, as pointed out by Sonesson and Bock (2003).

4.2 *Likelihood based methods*

The methods differ with respect to in what way the partial likelihood ratios

$$L(s, t) = f_{X_s}(x_s | \tau=t) / f_{X_s}(x_s | D), \quad t=1, \dots, s,$$

for a change at $\tau=t$, are weighted. The methods depend on different parameters which can be chosen to make them optimal for specific situations such as one with an intensity v and shift size d .

The method based on the full likelihood ratio, the LR method, has the alarm statistic

$$p(x_s) = f_{X_s}(x_s | C(s)) / f_{X_s}(x_s | D(s)) = \sum_{t=1}^s w(t) \cdot L(s, t)$$

where $w(t) = P(\tau=t) / P(\tau \leq s)$ is the weight for $L(s, t)$. It was shown by Frisén and de Maré (1991) that the alarm rule of the LR method can be expressed in terms of the posterior probability $P(C(s) | x_s)$ and a positive constant limit g_{PP} . This is equivalent to the LR method with the limit $g_{LR}(s) = g_{PP} / (1 - g_{PP}) \cdot P(D(s)) / P(C(s)) = g_{PP} / (1 - g_{PP}) \cdot P(\tau > s) / P(\tau \leq s)$. The LR method depends on the specified v and d , for which the method is optimized. For a geometric distribution the method is ED optimal for a process with the parameter values used.

A likelihood ratio method based on a small intensity ($v \rightarrow 0$) is the SR method (Shiryayev (1963) and Roberts (1966)) which can be used when the distribution of τ is unknown. From a Bayesian point of view this method can be seen as based on a non-informative generalized prior for τ since the weights $w(t)$ tends to a constant. Also the alarm limit $g(s)$ tend to a constant. The SR method depends only on d and can be used as an approximation to the LR method. Frisén and Wessman (1999) showed that the approximation works well, even for as large intensities as $v=0.20$.

The CUSUM method of Page (1954) uses $p(x_s) = \max_{1 \leq t \leq s} \{L(s, t)\}$ where $g(s)$ is a constant g . It depends on d and satisfies the minimax criterion described in the previous section. The Shewhart method uses $p(x_s) = L(s, s)$ and a constant limit, i.e. an alarm is given as soon as the last observations exceeds the limit. It has no dependency on v and d . The Shewhart method is ED optimal when $C(s) = \{\tau=s\}$ and $D(s) = \{\tau > s\}$ because then the alarm statistic of the LR method reduces to $L(s, s)$.

For the situation specified in section 2 where $\mu=0$ the partial likelihood ratios are

$$L(s, t) = d^{-(s-(t-1))/2} \cdot \exp\left\{\delta(d, \sigma^2) \cdot \sum_{i=t}^s x^2(i)\right\}, \quad t=1, \dots, s$$

where $\delta(d, \sigma^2) = (2 \cdot \sigma^2)^{-1} \cdot (1 - d^{-1})$. The alarm statistic of the LR method can then be expressed as

$$p(x_s) = \left(d^{s/2} \cdot P(\tau \leq s)\right)^{-1} \cdot \sum_{t=1}^s P(\tau = t) \cdot d^{(t-1)/2} \cdot \exp\left\{\delta(d, \sigma^2) \cdot \sum_{i=t}^s x^2(i)\right\},$$

which can be written recursively as

$$p(x_s) = \left(P(\tau \leq s-1) / \left(d^{1/2} \cdot P(\tau \leq s)\right)\right) \cdot \left\{\exp\left\{\delta(d, \sigma^2) \cdot x^2(s)\right\} \cdot \left\{p(x_{s-1}) + P(\tau = s) / P(\tau \leq s-1)\right\}\right\},$$

$s=2, 3, \dots, p(x_1) = d^{-1/2} \cdot \exp\left\{\delta(d, \sigma^2) \cdot x^2(1)\right\}$. The SR method has the alarm statistic

$$p(x_s) = d^{-s/2} \cdot \sum_{i=1}^s d^{(i-1)/2} \cdot \exp\left\{\delta(d, \sigma^2) \cdot \sum_{i=1}^s x^2(i)\right\},$$

which can be written recursively as $p(x_s) = d^{-1/2} \cdot \left\{ \exp\left\{\delta(d, \sigma^2) \cdot x^2(s)\right\} \cdot \left\{p(x_{s-1}) + 1\right\}\right\}$, $s=2, 3, \dots$, $p(x_1) = d^{-1/2} \cdot \exp\left\{\delta(d, \sigma^2) \cdot x^2(1)\right\}$. The alarm statistic of the CUSUM method can be written recursively as

$$p(x_s) = \max\left\{0, p(x_{s-1}) + x^2(s) - k\right\}$$

where $p(x_0) = 0$ and $k = (2 \cdot \delta(d, \sigma^2))^{-1} \cdot \ln d$. It can be shown that $\sigma^2 \leq k \leq d \cdot \sigma^2$. The alarm rule of the Shewhart method can be written as

$$p(x_s) = x^2(s) / \sigma^2 > g.$$

4.3 Limiting equalities

It was proven by Frisén and Wessman (1999) that when the size of the change in the mean for which the methods are optimized for tends to infinity the stopping rules of LR, SR and CUSUM tends to the stopping rule of the Shewhart method. Below we prove the same behavior for a change in the variance.

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

Proof:

$$\begin{aligned} p(x_s) > g_{LR}(s) &\Leftrightarrow \left(d^{s/2} \cdot P(\tau \leq s)\right)^{-1} \cdot \sum_{i=1}^s P(\tau = i) \cdot d^{(i-1)/2} \cdot \exp\left\{\delta(d, \sigma^2) \cdot \sum_{i=1}^s x^2(i)\right\} > \\ &g_{PP} / (1 - g_{PP}) \cdot P(\tau > s) / P(\tau \leq s) \Leftrightarrow \\ &\sum_{i=1}^s P(\tau = i) \cdot d^{(i-1)/2} \cdot \exp\left\{\delta(d, \sigma^2) \cdot \sum_{i=1}^s x^2(i)\right\} > (g_{PP} / (1 - g_{PP})) \cdot d^{s/2} \cdot P(\tau > s) \Leftrightarrow \\ &\exp\left\{\delta(d, \sigma^2) \cdot x^2(s)\right\} \cdot \left[\sum_{i=1}^{s-1} P(\tau = i) \cdot d^{(i-1)/2} \cdot \exp\left\{\delta(d, \sigma^2) \cdot \sum_{i=1}^{s-1} x^2(i)\right\} + P(\tau = s) \cdot d^{(s-1)/2}\right] \\ &> (g_{PP} / (1 - g_{PP})) \cdot d^{s/2} \cdot P(\tau > s) \Leftrightarrow \exp\left\{\delta(d, \sigma^2) \cdot x^2(s)\right\} > \\ &\frac{g_{PP} \cdot d^{s/2} \cdot P(\tau > s)}{(1 - g_{PP}) \cdot \sum_{i=1}^{s-1} P(\tau = i) \cdot d^{(i-1)/2} \cdot \exp\left\{\delta(d, \sigma^2) \cdot \sum_{i=1}^{s-1} x^2(i)\right\} + P(\tau = s) \cdot d^{(s-1)/2}} \Leftrightarrow \\ &\exp\left\{\delta(d, \sigma^2) \cdot x^2(s)\right\} > \\ &\frac{g_{PP}}{(1 - g_{PP}) \cdot v \cdot \sum_{i=1}^{s-1} (1 - v)^{-(s-t+1)} \cdot d^{-(s-t+1)/2} \cdot \exp\left\{\delta(d, \sigma^2) \cdot \sum_{i=1}^{s-1} x^2(i)\right\} + (1 - v)^{-1} \cdot d^{-1/2}} \Leftrightarrow \\ &x^2(s) > \frac{\ln g_{PP}}{\delta(d, \sigma^2)} - \frac{1}{\delta(d, \sigma^2)} \times \\ &\ln\left\{(1 - g_{PP}) \cdot v \cdot \sum_{i=1}^{s-1} \left((1 - v)^{-(s-t+1)} \cdot d^{-(s-t+1)/2} \cdot \exp\left\{\delta(d, \sigma^2) \cdot \sum_{i=1}^{s-1} x^2(i)\right\}\right) + (1 - v)^{-1} \cdot d^{-1/2}\right\} \end{aligned}$$

$$\Leftrightarrow x^2(s) > \frac{\ln g_{pp}}{\delta(d, \sigma^2)} - \frac{1}{\delta(d, \sigma^2)} \cdot \ln \left\{ (1-g_{pp}) \cdot v \cdot O(d^{-1}) + (1-v)^{-1} \cdot d^{-1/2} \right\}.$$

The dependency on s of the right hand side of the last expression disappears when d tends to infinity such that the stopping rule tends to the one of the Shewhart method.

Theorem 2: The stopping rule of the SR method tends to that of the Shewhart method when d tends to infinity.

Proof: In analogy with the proof of Theorem 1.

Theorem 3: The stopping rule of the CUSUM method tends to that of the Shewhart method when d tends to infinity.

Proof: $d \rightarrow \infty \Rightarrow k = (2 \cdot \delta(d, \sigma^2))^{-1} \cdot \ln d \rightarrow \infty \Rightarrow P(\max\{0, p(x_{s-1}) + x^2(s) - k\} > g) \rightarrow P(x^2(s) - k > g) = P(x^2(s) > g + k)$ since $\lim_{k \rightarrow \infty} P(p(x_{s-1}) > 0) = 0$.

5 A MONTE CARLO STUDY

In this section, we study the properties of the methods. To make the methods comparable the alarm limits are adjusted to yield the same level of MRL^0 . Which level of MRL^0 that should be chosen and what size of the scale change to be studied depends on the application.

A low value of MRL^0 can be interpreted as a situation where observations are made seldom and a high value with more frequent observations. This can be interpreted as differences in time scale. How distinct the differences between the methods are depends on the scale, as pointed out by Frisén and Wessman (1999). For example, if observations are made frequent, e.g. each day (a large value of MRL^0) then there is a larger information loss of only using the last observation (Shewhart method) compared to less frequent observations, e.g. each week (a small value of MRL^0). Comparisons with different values of MRL^0 are not made here. The alarm limits are set here such that $MRL^0 = 60$ which reflects roughly three months of daily data in the financial markets.

The in-control and out-of-control variance is set to 1 and 2, respectively, i.e. $\sigma^2 = 1$ and $\Delta = 2$ in section 2, as in e.g. MacGregor and Harris (1993) and Acosta-Mejia et al. (1999). The size of the change for which the methods are optimized for, d , is set to 1.5, 2 and 2.5. For $d = 2$ the variance is correctly specified whereas for d equals to 1.5 and 2.5 the variance is under- and over specified by 50%, respectively. The value of the intensity for which the LR method is optimized for, v , is set to 0.10 and 0.20. To distinguish between the same methods with different values of v and d , the values will be given as arguments, e.g. LR(v ; d).

For the Shewhart method analytical calculations were made. For the other methods simulations of 10^7 replicates were made. The limits were set such that the largest deviation between the values of $P(t_A \leq 60 | D)$ and the intended values of 0.50 were smaller than 0.1%.

5.1 In-control properties

Having equal MRL^0 do not mean that the in-control run length densities are identical but they can have different shapes. The most common way to control the false alarms is by the ARL^0 . $MRL^0=60$ corresponds to values of ARL^0 between 60 and 87 for the methods. The great variation in ARL^0 is due to the great differences in skewness seen in the in-control run length densities shown in Figure 1. LR(0.2) yield the smallest values of ARL^0 for all values of d and Shewhart the largest and these two methods have densities that are most symmetric and skewed, respectively. Shewhart and CUSUM have similar ARL^0 . As implied by the theorems, the larger the d the more similar are the densities to the one of the Shewhart method. A method designed to detect a large change quickly should allocate nearly all weight to the single last observation, as pointed out by Frisén and Wessman (1999). When the methods are optimized for detecting a small change in the variance, many observations are required to have enough evidence for a change and the densities are consequently less skewed compared to when d is large.

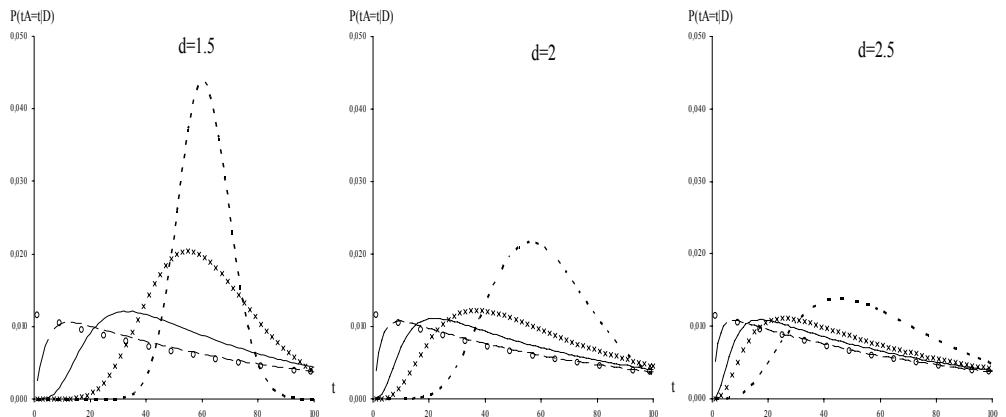


Figure 1. The density of the time of alarm, $P(t_A=t|D)$. Shewhart(\circ), CUSUM($---$), SR($—$), LR(0.1)(\times), LR(0.2)($---$).

It seems surprising that LR(0.2) is less skewed compared to LR(0.1) and SR since a large intensity should intuitively yield a large probability of early alarms. This was also noted by Frisén and Wessman (1999) who explained it by the way the false alarms are controlled. For a low intensity the right-hand tail of the run length distribution is thick. As ARL^0 was fixed, the only possibility was a high alarm probability at early times. When MRL^0 is fixed the time points of the alarms have less effect but many alarm times larger than 60 must still be compensated by high alarm probabilities early.

The probability of a false alarm, PFA, is another measure used to control the false alarms. It summarizes the false alarm distribution by weights with the distribution of τ . It is shown as a function of v in Figure 2.

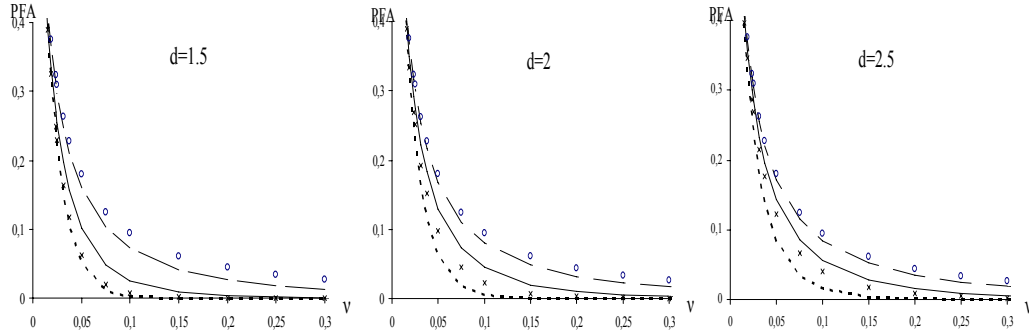


Figure 2. The probability of a false alarm, PFA. Shewhart(\circ), CUSUM(---), SR(—), LR(0.1)(\times), LR(0.2)(- - - -).

The differences in PFA are due to both the differences in shape of $P(t_A < \tau | D)$ and the location. As a result of the shape of the geometric distribution early alarms have a great influence on PFA. The large PFA for the Shewhart method is a result of the many early false alarms seen in Figure 1. Due to the opposite behaviour of the error spending of LR(0.2) it has the smallest PFA. Like an equal MRL^0 does apparently not imply equal PFA and vice versa, Frisén and Wessman (1999) and Frisén and Sonesson (2003) demonstrated the same difference between ARL^0 and PFA. Consequently, comparisons between methods depend on which measure that is controlled.

5.2 Out-of-control properties

As was mentioned in section 3 the out-of-control behaviour is in often summarized by the ARL^1 . In

Figure 3 the ARL^1 is shown as a function of d . The convergence to the ARL^1 of the Shewhart method is evident.

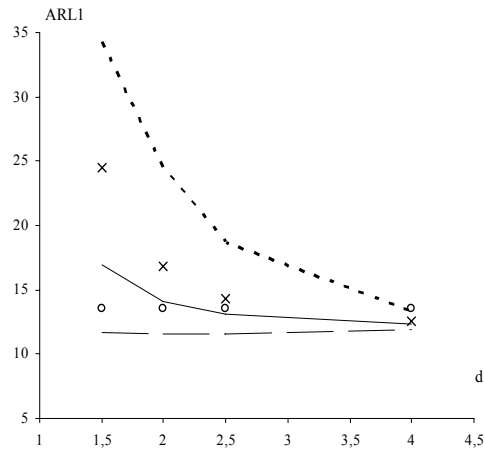


Figure 3. ARL^1 as function of d . $\Delta=2$. Shewhart(\circ), CUSUM(---), SR(—), LR(0.1)(\times), LR(0.2)(- - - -).

The conditional expected delay, CED, is shown in Figure 4 for different values of τ . For $\tau=1$, $CED=ARL^1-1$. The CED clearly depends on τ for several methods, which is not revealed by the ARL^1 . The worst value of CED is at $t=1$ for the CUSUM method

and CUSUM has the smallest CED(1) among the methods. Though ARL^0 is not controlled here this illustrates the minimax optimality of the CUSUM method.

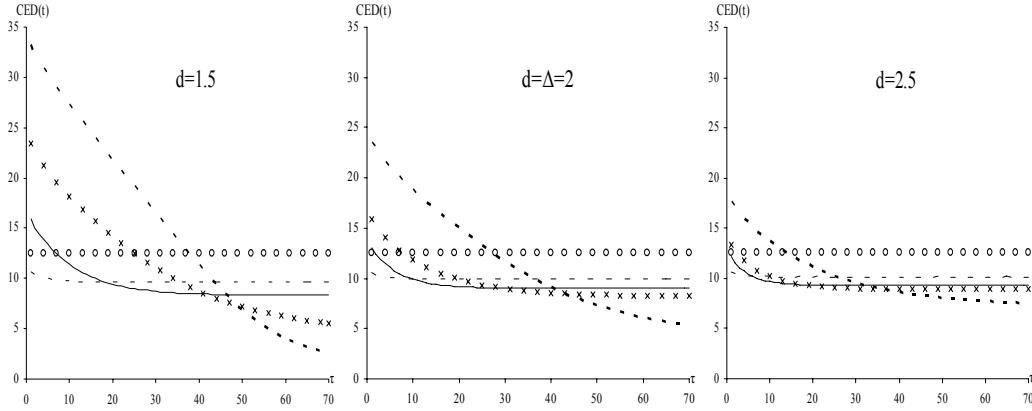


Figure 4. Conditional expected delay, $CED(\tau)$. Shewhart(\circ), CUSUM($---$), SR($—$), LR(0.1)(\times), LR(0.2)($-.-.-$).

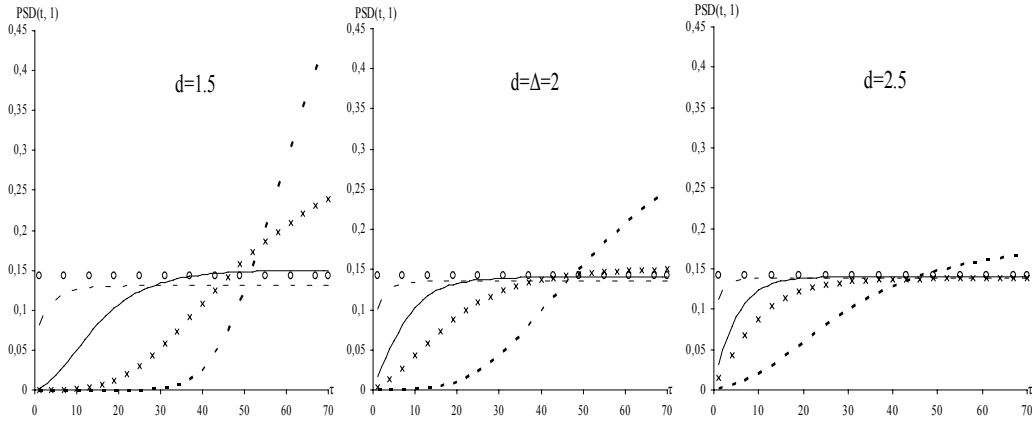


Figure 5. Probability of successful detection, $PSD(m, \tau)$, with $m=1$. Shewhart(\circ), CUSUM($---$), SR($—$), LR(0.1)(\times), LR(0.2)($-.-.-$).

CUSUM is better in terms of CED but worse in terms of PSD with $m=1$ (Figure 5) compared to Shewhart. The reason for this is that $P(t_A=t | t_A \geq \tau, \tau=t)$ that is PSD with $m=0$, is higher for Shewhart compared to CUSUM which favors PSD with $m=1$. The high $P(t_A=t | t_A \geq \tau, \tau=t)$ of the Shewhart method is due to its optimality for $C(s)=\{\tau=s\}$ and $D(s)=\{\tau > s\}$ (see section 4.2). The error spending behavior of the LR methods explained in section 5.1 influences the detection ability such that the methods have a large CED and a small PSD for early changes and the opposite for late τ .

5.3 The trust of alarms

The predictive value at time t , $PV(t)$, reflects the trust you should have in an alarm. The predictive value at time point t is

$$P(\tau \leq t | t_A=t) = \frac{PMA(t)}{PMA(t) + PFA(t)}$$

where $PFA(t)=P(t_A=t|t < \tau) \cdot P(\tau > t)$ and $PMA(t)=\sum_{i=1}^t P(\tau=i) \cdot P(t_A=t|\tau=i)$ are probabilities of a false and a motivated alarm at time t , respectively. The PV is shown as a function of the time of the alarm in Figure 6 and 7.

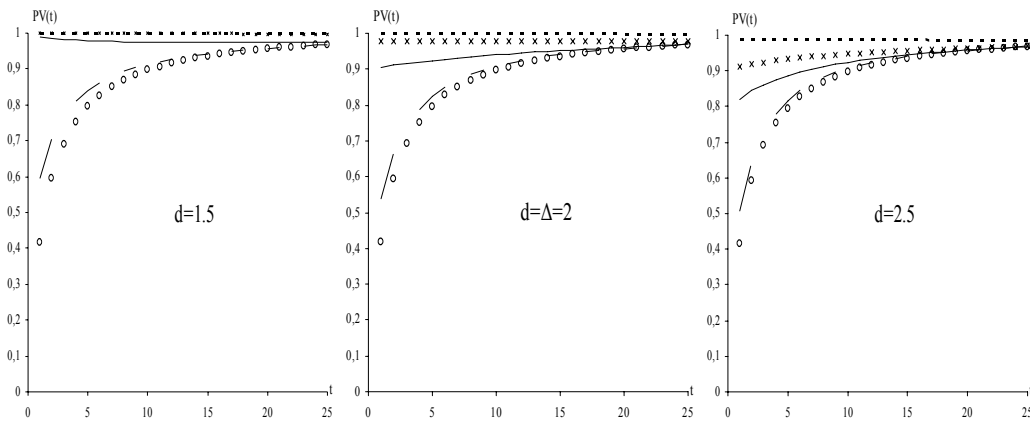


Figure 6. Predictive value, PV, with $v=0.10$. Shewhart(\circ), CUSUM($---$), SR($---$), LR(0.1)(\times), LR(0.2)($---$).

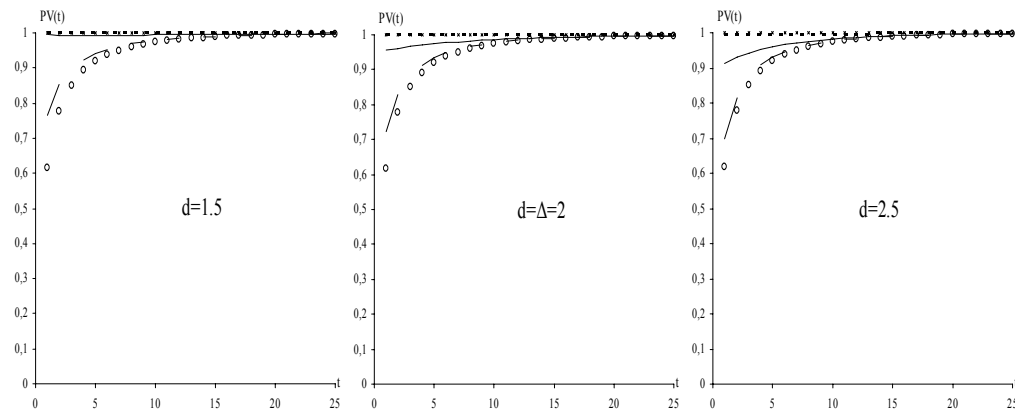


Figure 7. Predictive value, PV, with $v=0.20$. Shewhart(\circ), CUSUM($---$), SR($---$), LR(0.1)(\times), LR(0.2)($---$).

Shewhart and CUSUM have high detection ability for early changes as seen in Figure 3, 4, and 5, but at the same time a high false alarm probability (Figure 1). The result is a low predictive value of early alarms, i.e. these are not very trustworthy. The results get better for a large value of v . The PV of SR and LR appear to be fairly robust to mis-specifications of Δ . For these methods PV is stable over time, which might be a desirable property as it simplifies matters if the same action can be used regardless of whenever an alarm occurs.

6 ILLUSTRATIVE EXAMPLE

The use of the methods is here illustrated by a simple example. The methods are monitoring the returns, denoted by r_t , of the stock market index Standard and Poor's 500 (S&P500).

Andreou and Ghysels (2002) applied a number of tests for homogeneity in the variance of the returns of S&P500 for the period 4 January 1989-19 October 2001 (3229 observations). The tests were made retrospectively that is a historical data set of given length were analyzed. Both tests for a single change point and for multiple change points were applied. For the latter the number of breaks was determined by the test of Kokoszka and Leipus (2000) applied to the squared returns and a sequential segmentation approach. It was concluded that changes in the volatility occurred at 31 December 1991, 18 December 1995 and 26 March 1997. Whether the second change in 18 December 1995 could have been detected online is investigated below.

The period of monitoring is 9 October 1995–25 March 1997 that is 370 observations and $\tau=50$. Financial returns are known to be conditionally heteroscedastic. We try to explain the heteroscedasticity by an in-control model and monitor the residuals. An in-control model is estimated by a historical set of data 31 December 1991 – 6 October 1995 (954 observations). The returns of the historical period and the monitoring period are shown in Figure 8.

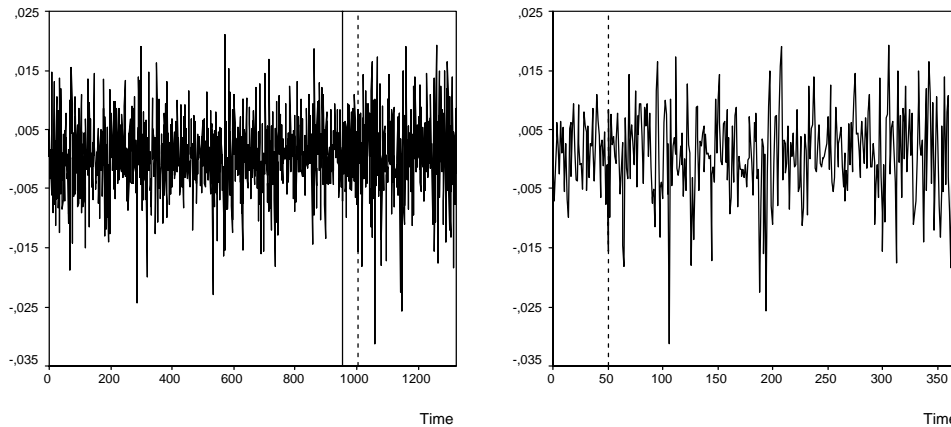


Figure 8. Daily returns of S&P500. The start of the monitoring period is marked with a solid vertical line. The time of the change τ is marked with a dashed vertical line. Left: 31 December 1991 – 25 March 1997. Right: 9 October 1995 – 25 March 1997.

A common way of characterizing the heteroscedasticity is by ARCH processes. The portmanteau Q-test of the squared residuals (McLeod and Li (1983)) and the Lagrange multiplier test by Engle (1982) for ARCH disturbances are applied to the returns of the historical period. The tests can be used to identify the order of an ARCH process. The p-values at different lags are shown in Table 1 and these indicate that there are high order ARCH effects which could be described by a first order Gaussian GARCH process, $\text{GARCH}(1, 1)$; $r(t)=\mu(t)+\varepsilon(t)\cdot h(t)^{1/2}$ where $h(t)=\omega+\alpha_1\cdot(r(t-1)-\mu(t-1))^2+\beta_1\cdot h(t-1)$, $\omega>0$, $\alpha_1>0$, $\beta_1\geq 0$, $\alpha_1+\beta_1<1$ and $\varepsilon\sim\text{iid } N(0, 1)$. We estimate the parameters of the Gaussian $\text{GARCH}(1, 1)$ model with $\mu(t)=\mu$ (a constant) using the historical data set. The parameter estimates are given in Table 1. It should be pointed out that a proper modeling strategy requires a much more thorough analysis than this. But for illustration the model is used as a rough approximation.

The statistic under surveillance is the squares of the standardized residuals $X(t)=(r(t)-\hat{\mu})/\hat{h}(t)^{1/2}$ where $\hat{h}(t)$ is the estimated conditional variance.

Since Δ is unknown the values of d , for which the methods are optimized for, used in section 5 are used also here. The alarm limits used earlier are used also here. The alarm times are given in Table 1.

P-values of the portmanteau Q-test and the Lagrange multiplier test at different lags.

Lag	1	2	3	4	5	6	7	8
Q-test	0.442	0.670	0.233	0.087	0.044	0.025	0.036	0.054
LM-test	0.440	0.671	0.237	0.099	0.058	0.041	0.049	0.077

Parameter estimates of the GARCH(1, 1) model. Standard errors within brackets.

μ	ω	α_1	β_1
0.000452 (0.000183)	1.6561E-6 (7.8792E-7)	0.0356 (0.0125)	0.9134 (0.0327)

Alarm times ($\tau=50$).

LR(1.5; 0.1) 51	LR(2; 0.1) 50	LR(2.5; 0.1) 50
LR(1.5; 0.2) 55	LR(2; 0.2) 51	LR(2.5; 0.2) 50
SR(1.5) 50	SR(2) 50	SR(2.5) 50
CUSUM(1.5) 51	CUSUM(2) 50	CUSUM(2.5) 50
Shewhart 50		

Table 1. Results from the modeling strategy and the alarm times of the methods.

All the methods give alarms at τ or immediately after. The variances of X before and after τ as estimated by $\hat{\sigma}^2=(1/(n-1)) \cdot \sum_{t=1}^n (x(t)-\bar{x})^2$ yields a shift of the size $\hat{\Delta}=1.495$. At τ there is a highly negative return (see Figure 8, right) influencing the alarm statistics. For the model used in the simulations $P(t_A=\tau|\tau=50)$ varies between 0.035 and 0.057 for the methods and the outcome in Table 1 is hence rather extreme. The residuals thus appear to deviate from the process of interest. The validity of the Gaussian GARCH(1, 1) model for describing financial time series is in fact frequently debated in the empirical finance literature. This illustrates many of the difficulties encountered by case studies.

7 CONCLUDING REMARKS

Different likelihood based methods of statistical surveillance for detecting a change in the variance has been evaluated. The methods differ with respect to how the different observations available at each decision time are treated and the way the alarm limit change with the decision time. The methods differ with respect to the number of parameters they depend on. All methods but the Shewhart method depends on the size of the change and the LR methods does also depend on the intensity of the change-point time.

The robustness of the methods with respect to mis-specifications of the change has been examined. The results demonstrates the same behavior Frisén and Wessman (1999) found for a change in location: the larger the size of the change, d , for which

the methods are optimized for, the more similar to the Shewhart method are the methods. Hence, if we optimize for a very large d all weight is allocated to the last observation. If, on the other hand, we optimize for a small d , more weight is given to earlier observations than for Shewhart because more observations are needed for having enough evidence for a change.

Differences in the weighting are reflected in the skewness of the run length densities. For a large d early alarms are more frequent compared to a small d . Consequences of these differences are that for the former situation PFA is high compared to the latter as early alarms have a great influence because of the geometric distribution.

The detection ability as measured by CED and PSD is rather constant and great at a large d . When the d is small the detection ability is worse at early changes but get better the later the change occurs. The price of the good detection ability of early changes of Shewhart and CUSUM is however that early alarms are not very reliable. LR and SR have better predictive values at early alarms.

The LR method has the parameter v to optimize for the intensity for a change. This is avoided by the SR method. For the values of MRL^0 and v used, LR seems in terms of PFA and PV (Figure 2, 6 and 7) to be robust against mis-specifications of the intensity and SR appear to be a good approximation of LR for small values of the intensity.

The surprising way the methods differed in shapes of the run length densities noted by Frisén and Wessman (1999) is also seen here. It depends on the way the false alarms are controlled, as explained in section 5.1.

In the illustration of the methods on the S&P500 data all the methods gave alarms very close to the change-point time. However, if the model residuals represented the process of interest then these results would be very improbable. This illustrates many of the difficulties and limitations encountered by case studies.

ACKNOWLEDGEMENTS

The author is grateful for valuable comments by Professor Marianne Frisén. The Bank of Sweden Tercentenary Foundation supported the research.

REFERENCES

- Acosta-Mejia, C. A. (1998) Monitoring reduction in variability with the range. *IIE Transactions*, **30**, 515-523.
- Acosta-Mejia, C. A., Pignatello, J. J. J. and Rao, B. V. (1999) A comparison of control charting procedures for monitoring process dispersion. *IIE Transactions*, **31**, 569-579.
- Acosta-Mejia, C. A. and Pignatiello, J. J. (2000) Monitoring process dispersion without subgrouping. *Journal of Quality Technology*, **32**, 89-102.
- Andersson, E., Bock, D. and Frisén, M. (2004) Detection of turning points in business cycles. *Journal of business cycle measurement and analysis*, **1**, 93-108.

- Andersson, E., Bock, D. and Frisé, M. (2005) Statistical surveillance of cyclical processes. Detection of turning points in business cycles. *Journal of Forecasting*, **24**, 465-490.
- Andersson, E., Bock, D. and Frisé, M. (2006) Some statistical aspects on methods for detection of turning points in business cycles. *Journal of Applied Statistics*, **33**, 257-278.
- Andreou, E. and Ghysels, E. (2002) Detecting multiple breaks in financial market volatility dynamics. *Journal of Applied Econometrics*, **17**, 579-600.
- Bock, D. (2006) Aspects on the control of false alarms in statistical surveillance and the impact on the return of financial decision systems. *Submitted*.
- Bock, D., Andersson, E. and Frisé, M. (2006) The relation between statistical surveillance and certain decision rules in finance. *Submitted*.
- Chen, G., Cheng, S. W. and Xie, H. (2004) A New EWMA Control Chart for Monitoring Both Location and Dispersion. *Quality Technology & Quantitative Management*, **1**, 217-231.
- Chu, C.-S. J., Stinchcombe, M. and White, H. (1996) Monitoring structural change. *Econometrica*, **64**, 1045-1065.
- Costa, A. F. B. and Rahim, M. A. (2004) Monitoring Process Mean and Variability with One Non-central Chi-square Chart. *Journal of Applied Statistics*, **31**, 1171 - 1183.
- Crowder, S. V. and Hamilton, M. D. (1992) An EWMA for monitoring a process standard-deviation. *Journal of Quality Technology*, **24**, 12-21.
- Domangue, R. and Patch, S. C. (1991) Some omnibus exponentially weighted moving average statistical process monitoring schemes. *Technometrics*, **33**, 299-313.
- Engle, R. F. (1982) Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica*, **50**, 987-1008.
- Frisé, M. (1992) Evaluations of Methods for Statistical Surveillance. *Statistics in Medicine*, **11**, 1489-1502.
- Frisé, M. (1994) Statistical Surveillance of Business Cycles. Research Report, Department of Statistics, Göteborg University, Sweden, 1994:1 (Revised 2000)
- Frisé, M. (2003) Statistical Surveillance. Optimality and Methods. *International Statistical Review*, **71**, 403-434.
- Frisé, M. and de Maré, J. (1991) Optimal Surveillance. *Biometrika*, **78**, 271-80.
- Frisé, M. and Sonesson, C. (2003) Optimal surveillance based on exponentially weighted moving averages methods. *Submitted*.
- Frisé, 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. (1993) An optimal-design of EWMA control charts based on median run-length. *Journal of Statistical Computation and Simulation*, **45**, 169-184.
- Hawkins, D. L. (1992) Detecting shifts in functions of multivariate location and covariance parameters. *Journal of Statistical Planning and Inference*, **33**, 233-244.
- Hawkins, D. M. (1981) A CUSUM for a scale parameter. *Journal of Quality Technology*, **13**, 228-231.
- Hawkins, D. M. and Zamba, K. D. (2005) A change point model for a shift in the variance. *Journal of Quality Technology*, **37**, 21-37.
- Järpe, E. and Wessman, P. (2000) Some power aspects of methods for detecting shifts in the mean. *Communications in Statistics. Simulations and Computations*, **29**.

- Kokoszka, P. and Leipus, R. (2000) Change-point estimation in ARCH models. *Bernoulli*, **6**, 513-539.
- Lai, T. L. (1995) Sequential Changepoint Detection in Quality-Control and Dynamical-Systems. *Journal of the Royal Statistical Society B*, **57**, 613-658.
- MacGregor, J. F. and Harris, T. J. (1993) The exponentially weighted moving variance. *Journal of Quality Technology*, **25**, 106-118.
- McLeod, A. I. and Li, W. K. (1983) Diagnostic checking ARMA time series models using squared-residual autocorrelations. *Journal of Time Series Analysis*, **4**, 269-273.
- Moustakides, G. V. (1986) Optimal stopping times for detecting changes in distributions. *The Annals of Statistics*, **14**, 1379-1387.
- Ncube, M. and Li, K. (1999) An ewma-cuscore quality control procedure for process variability. *Mathematical and Computer Modelling*, **29**, 73-79.
- Page, E. S. (1954) Continuous inspection schemes. *Biometrika*, **41**, 100-114.
- Page, E. S. (1963) Controlling the standard deviation by Cusums and warning lines. *Technometrics*, **5**, 307-315.
- Petzold, M., Sonesson, C., Bergman, E. and Kieler, H. (2003) Surveillance in longitudinal models. Detection of intra-uterine growth restriction. *Biometrics*, **60**, 1025-1033.
- Pollak, M. and Siegmund, D. (1975) Approximations to the Expected Sample Size of Certain Sequential Tests. *Annals of Statistics*, **3**, 1267-1282.
- Reynolds, M. R. and Soumbos, Z. G. (2001) Monitoring the process mean and variance using individual observations and variable sampling intervals. *Journal of Quality Technology*, **33**, 181-205.
- Rigdon, S. E., Cruthis, E. N. and Champ, C. W. (1994) Design strategies for individuals and moving range control charts. *Journal of Quality Technology*, **26**, 274-287.
- Roberts, S. W. (1959) Control Chart Tests Based on Geometric Moving Averages. *Technometrics*, **1**, 239-250.
- Roberts, S. W. (1966) A Comparison of some Control Chart Procedures. *Technometrics*, **8**, 411-430.
- Schipper, S. and Schmid, W. (2001a) Control charts for GARCH processes. *Nonlinear Analysis*, **47**, 2049-2060.
- Schipper, S. and Schmid, W. (2001b) Sequential Methods for Detecting Changes in the Variance of Economic Time Series. *Sequential Analysis*, **20**, 235-262.
- Severin, T. and Schmid, W. (1998) Statistical process control and its application in finance. In *Risk measurement, econometrics and neural networks*(Eds, Bol, G., Nakhaeizadeh, G. and Vollmer, C.-H.) Physica Verlag, Heidelberg, pp. 83-104.
- Severin, T. and Schmid, W. (1999) Monitoring changes in GARCH processes. *Allgemeines Statistisches Archiv*, **83**, 281-307.
- Shiryayev, A. N. (1963) On optimum methods in quickest detection problems. *Theory of Probability and its Applications.*, **8**, 22-46.
- Shiryayev, A. N. (2002) Quickest Detection Problems in the Technical Analysis of Financial Data. In *Mathematical Finance - Bachelier Congress 2000*(Eds, Geman, H., Madan, D., Pliska, S. and Vorst, T.) Springer.
- Sonesson, C. and Bock, D. (2003) A review and discussion of prospective statistical surveillance in public health. *Journal of the Royal Statistical Society A*, **166**, 5-21.

- Srivastava, M. (1997) Cusum procedures for monitoring variability. *Communications in Statistics. Theory and Methods*, **26**, 2905-2926.
- Srivastava, M. S. and Chow, W. (1992) Comparison of the CUSUM procedure with other procedures that detect an increase in the variance and a fast accurate approximation for the ARL of the CUSUM procedure. Research Report, Department of Statistics, University of Toronto, 9122
- Srivastava, M. S. and Wu, Y. (1993) Comparison of EWMA, CUSUM and Shiriyayev-Roberts Procedures for Detecting a Shift in the Mean. *Annals of Statistics*, **21**.
- Wetherhill, G. B. and Brown, D. W. (1991) *Statistical Process Control: Theory and Practice*, Chapman and Hill, London.
- Yashchin, E. (1993) Statistical Control Schemes - Methods, Applications and Generalizations. *International Statistical Review*, **61**, 41-66.

Research Report

- | | | |
|--------|---|--|
| 2007:1 | Andersson, E.: | Effect of dependency in systems for multivariate surveillance. |
| 2007:2 | Frisén, M.: | Optimal Sequential Surveillance for Finance, Public Health and other areas. |
| 2007:3 | Bock, D.: | Consequences of using the probability of a false alarm as the false alarm measure. |
| 2007:4 | Frisén, M.: | Principles for Multivariate Surveillance. |
| 2007:5 | Andersson, E., Bock, D. & Frisé, M.: | Modeling influenza incidence for the purpose of on-line monitoring. |
| 2007:6 | Bock, D., Andersson, E. & Frisé, M.: | Statistical Surveillance of Epidemics: Peak Detection of Influenza in Sweden. |
| 2007:7 | Andersson, E., Kühmann-Berenzon, S., Linde, A., Schiöler, L., Rubinova, S. & Frisé, M.: | Predictions by early indicators of the time and height of yearly influenza outbreaks in Sweden |
| 2007:8 | Bock, D., Andersson, E. & Frisé, M.: | Similarities and differences between statistical surveillance and certain decision rules in finance. |