

1

Introduction to financial surveillance

Marianne Frisé

Statistical Research Unit, School of Business, Economics and Law, Göteborg University, PO Box 660, SE 405 30 Göteborg, Sweden

1.1 What is financial surveillance?

In financial surveillance the aim is to signal at the optimal trading time. A systematic decision strategy is used. The information available at each possible decision time is evaluated in order to judge whether or not there is enough information for a decision about an action or if more information is necessary so that the decision should be postponed. Financial surveillance gives timely decisions.

The authors of this book hope that it will serve two purposes. First, we hope that it will stimulate an increased use of surveillance in finance by providing methods which have not been available before. Second, we hope that the statistical community will use the book as a spur to further development of techniques and to research on questions unanswered by the following chapters. Financial surveillance is a new area, and some open problems are discussed in Chapter 9.

Financial decision strategies are based, in one way or another, on continuous observation and analysis of information. This is financial surveillance. Statistical surveillance uses decision theory and statistical inference in order to derive timely decision strategies. Hopefully, this book will serve as a bridge between finance and statistical surveillance.

This book is written by statisticians with an interest in finance. Textbooks describing financial problems and statistical methods are for example Föllmer and Schied (2002), Härdle, Kleinow and Stahl (2002), Gouriéroux and Jasiak (2002), Franke, Härdle and Hafner (2004), Cizek, Härdle and Weron (2005) and Scherer and Martin (2005). Many and various statistical techniques are described in these books.

In Section 1.2 statistical methods which are useful for financial decisions will be discussed. In Section 1.3 the area of statistical surveillance is described, and the characteristics of surveillance are compared to other areas in statistics. Evaluations in surveillance are described in Section 1.4. This is an important area, since the choice of evaluation measures will decide which methods are considered appropriate. General methods for aggregating information over time are described in Section 1.5. Special aspects of surveillance for financial decisions are discussed in Section 1.6.

The content of the book will be described in Section 1.7, which deals with the relation between the chapters. This section also provides reading guidelines.

1.2 Statistical methods for financial decision strategies

Statistical methods use observations of financial data to give information about the financial process which produces the data. This is in contrast to probability theory, where assumptions about the financial process are used to derive which observations will be generated.

1.2.1 Transaction strategies based on financial data

In finance, the relation between observations and decisions is often informal. Statisticians have taken on the role of presenting statistical summaries of quantitative data. In many areas, including finance, this means providing point and interval estimates for the quantities of interest. Methods for providing such summaries are highly formalized and constantly evolving. The discipline of statistics uses observations to make deductions about the real world. It has its own set of axioms and theorems besides those of probability theory. While decision making is the incentive for much statistical analysis, the process that transforms statistical summaries into decisions usually remains informal and ad hoc.

In finance the timeliness of transactions is important to yield a large return and a low risk. The concept of an efficient arbitrage-free market is of great interest. One central question is whether the history of the price of an asset contains information which can be used to increase the future return. A natural aim

is to maximize the return. The theory of stochastic finance has been based on an assumption of an efficient market where the financial markets are arbitrage-free and there is no point in trying to increase the return. Even though this view is generally accepted today, there are some doubts that it is generally applicable. When the information about the process is incomplete, as for example when a change point could occur, there may be an arbitrage opportunity, as demonstrated by Shiryaev (2002). In Chapter 3 it will be discussed how technical analysis relies on the possibility of using history to increase future returns. The support for the efficient market hypothesis depends on the knowledge about the model, as is discussed below in Section 1.2.2.

1.2.2 Modelling

In finance, advanced stochastic models are necessary to capture all empirical features. The expected value could depend on time in a complicated nonlinear way. Parameters other than the expected value are often of great interest, and the risk (measured by variance) is often of great concern. Complicated dependencies are common, which means that complicated measures of variance are necessary. Multivariate data streams are of interest for example when choosing portfolio. The models may be described in continuous or discrete time. Chapter 2 gives an overview of models of interest in finance. Statistical methods for estimation and model choice will also be briefly described in the chapter. The use of the models should be robust to errors in the model specification.

1.2.2.1 Stochastic model assumed known

When the stochastic model is assumed to be completely known there is no expected return to be gained. We will have an arbitrage-free market. We can use probability theory to calculate the optimal transaction conditions. Important contributions are found in the book by Shiryaev (1999) or in articles in the scientific journal *Finance and Stochastics*. Also the proceedings of the conference *Stochastic Finance* in 2004 and 2007 are informative on how to handle financial decisions when the model is completely known.

1.2.2.2 Incomplete knowledge about the stochastic model

When the model is not completely known, the efficient and arbitrage-free market assumptions are violated. Changes at unknown times are possible. One has to evaluate the information continuously to decide whether a transaction at that time is profitable. Statistical inference is needed for the decision (Shiryaev 2002).

1.2.3 Evaluation of information

Statistical inference theory gives guidelines on how to draw conclusions about the real world from data. Statistical hypothesis testing is suitable for testing a single hypothesis but not a decision strategy including repeated decisions, as will be further described in Section 1.4.1. Statistical surveillance is an important branch of inference. The relatively new area of statistical surveillance deals with the sequential evaluation of the amount of information at hand. It provides a theory for deciding at what time the amount of information is enough to make a decision and take action. This bridges the gap between statistical analysis and decisions.

In this book, we concentrate on the methodology of statistical surveillance. This methodology is of special interest for financial decision strategies, but it is also relatively new in finance. The ambition here is to give a comprehensive description of such aspects of statistical surveillance that may be of interest in finance. Thus the next sections of this chapter will give a short review on statistical surveillance.

1.3 What is statistical surveillance?

1.3.1 General description

Statistical surveillance means that a time series is observed with the aim of detecting an important change in the underlying process as soon as possible after the change has occurred. Statistical methods are necessary to separate important changes in the process from stochastic variation. The inferential problems involved are important for the applications and interesting from a theoretical viewpoint, since they bring different areas of statistical theory together.

Broad surveys and bibliographies on statistical surveillance are given by Lai, who concentrates on the minimax properties of stopping rules, by Woodall and Montgomery (1999) and Ryan (2000), who concentrate on control charts, and by Frisé (2003), who concentrates on the optimality properties of various methods.

The theory of statistical surveillance has developed independently in different statistical subcultures. Thus, the terminology is diverse. Different terms are used to refer to ‘statistical surveillance’ as described here. However, there are some differences in how the terms are used. ‘Optimal stopping rules’ are most often used in probability theory, especially in connection with financial problems. However, this does not always include the statistical inference from the observations to the model. Literature on ‘change-point problems’ does not always treat the case of continuous observations but often considers the case

of a retrospective analysis of a fixed number of observations. The term ‘early warning system’ is most often used in the economic literature. ‘Monitoring’ is most often used in medical literature and as a nonspecific term. Timeliness, which is important in surveillance, is considered in the vast literature on quality control charts, and here also the simplicity of procedures is stressed. The notations ‘statistical process control’ and ‘quality control’ are used in the literature on industrial production and sometimes also include other aspects than the statistical ones.

The statistical methods suitable for surveillance differ from the standard hypothesis testing methods. In the prospective surveillance situation, data accumulated over time is analysed repeatedly. A decision concerning whether, for example, the variance of the price of a stock has increased or not has to be made sequentially, based on the data collected so far. Each new possibility demands a new decision. Thus, there is no fixed data set but an increasing number of observations. In sequential analysis we have repeated decisions, but the hypotheses are fixed. In contrast, there are no fixed hypotheses in surveillance. The statistics derived for a fixed sample may be of great value also in the case of surveillance, but there are great differences between the systems for decision. The difference between hypotheses and on-line surveillance is best seen by studying the difference in evaluation measures (see Section 1.4.1).

In complicated surveillance problems, a stepwise reduction of the problem may be useful. Then, the statistics derived to be optimal for the fixed sample problem can be a component in the construction of the prospective surveillance system. This applies, for example, to the multivariate problems described in Section 1.6.7 in this chapter and in Chapters 5–7.

1.3.2 History

The first modern control charts were developed in the 1920s, by Walter A. Shewhart and coworkers at Bell Telephone Laboratories. In 1931 the famous book *Economic Control of Quality of Manufactured Product* (Shewhart 1931) was published. The same year Shewhart gave a presentation of the new technique to the Royal Statistical Society. This stimulated interest in the UK. The technique was used extensively during World War II both in the UK and in the US. In the 1950s, W. E. Deming introduced the technique in Japan. The success in Japan spurred the interest in the West, and further development started.

In the Shewhart method each observation is judged separately. The next important step was taken when Page suggested the CUSUM method for aggregating information over time. Shortly afterwards, Roberts (1959) suggested another method for aggregating information – the EWMA method. A method

based on likelihood which fulfils important optimality conditions was suggested by Shiryaev (1963).

In recent years there have been a growing number of papers in economics, medicine, environmental control and other areas, dealing with the need of methods for surveillance. The threat of bioterrorism and new contagious diseases has been an important reason behind the increased research activity in the theory of surveillance. Hopefully, the time is now ripe for finance to benefit from all these results.

1.3.3 Specifications of the statistical surveillance problem

The general situation of a change in distribution at a certain change-point time τ will now be specified. The variable under surveillance could be the observation itself or an estimator of a variance or some other derived statistic, depending on the specific situation. We denote the process by $X = \{X(t) : t = 1, 2, \dots\}$, where $X(t)$ is the observation made at time t . The purpose of the monitoring is to detect a possible change. The time for the change is denoted by τ (see Figure 1.1). This can be regarded either as a random variable or as a deterministic but unknown value, depending on what is most suitable for the application.

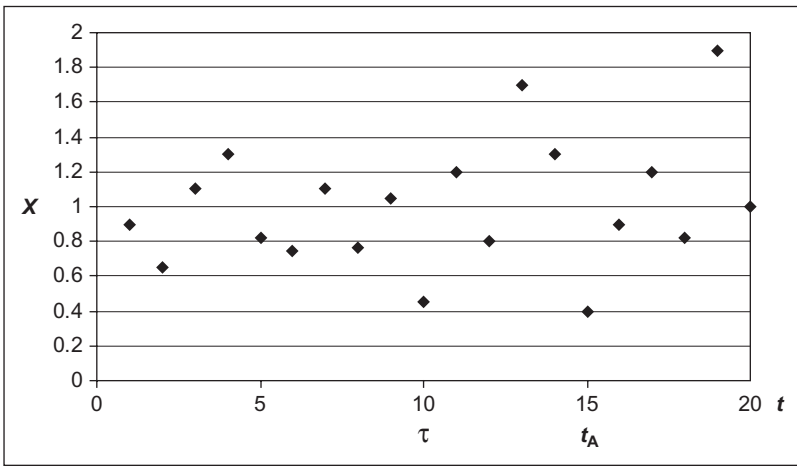


Figure 1.1 The first $\tau - 1$ observations $X_{\tau-1} = \{X(t) : t \leq \tau - 1\}$ are ‘in-control’ with a small variance. The subsequent observations (from $t = \tau$ (here 10) and onwards) have a larger variance. The alarm time is t_A , which happens to be 15. Thus the delay is $t_A - \tau = 5$.

The properties of the process change at time τ . In many cases we can describe this as

$$X(t) = \begin{cases} Y(t), & t < \tau \\ Y(t) + \Delta, & t \geq \tau \end{cases} \quad (1.1)$$

where Y is the ‘in-control’ or ‘target’ process and Δ denotes the change. More generally we can denote the ‘in-control’ state by D and the state which we want to detect by C . The (possibly random) process that determines the state of the system is here denoted by $\mu(t)$. This could be an expected value, a variance or some other time-dependent characteristic of the distribution. Different types of states between which the process changes are of interest for different applications. Descriptions of the details of the states are made within each chapter.

The change to be detected differs depending on the application. Most studies in literature concern a step change, where a parameter changes from one constant level, say, $\mu(t) = \mu^0$ to another constant level, $\mu(t) = \mu^1$. The case $\mu > 0$ is described here. We have $\mu(t) = \mu^0$ for $t = 1, \dots, \tau - 1$ and $\mu(t) = \mu^1$ for $t = \tau, \tau + 1$.

Even though autocorrelated time series are studied for example by Schmid and Schöne (1997), Petzold, Sonesson, Bergman and Kieler (2004), and in Chapters 5–7, processes which are independent given τ are the most studied and used also in Chapters 3 and 4. This simple situation will be used to introduce general concepts of evaluations, optimality and standard methods.

Some cases of special interest in financial surveillance are discussed in Section 1.6.

1.4 Evaluations

Quick detection and few false alarms are desired properties of methods for surveillance. Knowledge about the properties of the method in question is important. If a method calls an alarm, it is important to know whether this alarm is a strong indication of a change or just a weak indication. The same methods can be derived by Bayesian or frequentistic inference. However, evaluations differ. Here we present measures suitable for frequentistic inference.

1.4.1 The difference between evaluations for hypothesis testing and on-line surveillance

Measures for a fixed sample situation can be adopted for surveillance, but some important differences will be pointed out. In Table 1.1 the measures

Table 1.1 Evaluation measures for hypothesis testing and the corresponding measures for on-line surveillance.

	<i>Test</i>	<i>Surveillance</i>
<i>False alarms</i>	Size α , Specificity	ARL ⁰ , MRL ⁰ , PFA
<i>Detection ability</i>	Power, Sensitivity	ARL ¹ , MRL ¹ , CED, ED, maxCED, PSD, SADT

conventionally used in hypothesis testing and some measures for surveillance are given. These measures will be described and discussed below.

Different error rates and their implications for a decision system were discussed by Friséen and de Maré (1991). Using a constant probability of exceeding the alarm limit for each decision time means that we have a system of repeated significance tests. This may work well also as a system of surveillance and is often used. The Shewhart method described in Section 1.5.2 has this property. This is probably also the motive for using the limits with the exact variance in the EWMA method described in Section 1.5.4.

Evaluation by significance level, power, specificity and sensitivity, which is useful for a fixed sample, is not appropriate without modification in a surveillance situation since these measures do not have unique values in a surveillance system. One problem with evaluation measures originally suggested for the study of a fixed sample of, say, n observations is that the measures depend on n . For example, the specificity will tend to zero for most methods and the size of the test will tend to one when n increases.

Chu, Stinchcombe and White (1996) and others have suggested methods with a size less than one:

$$\lim_{n \rightarrow \infty} P(t_A \leq n|D) < 1.$$

This is convenient since ordinary statements of hypothesis testing can be made. However, Friséen (2003) demonstrated that the detection ability of methods with this property declines rapidly with the value of time τ of the change. Important consequences were illustrated by Bock (2008).

The performance of a method for surveillance depends on the time τ of the change. Generally, the sensitivity will not be the same for early changes as for late ones. It also depends on the length of time for which the evaluation is made. Thus, there is not one unique sensitivity value in surveillance, but other measures may be more useful. Accordingly, conventional measures for fixed samples should be supplemented by other measures designed for statistical surveillance, as will be discussed in the following.

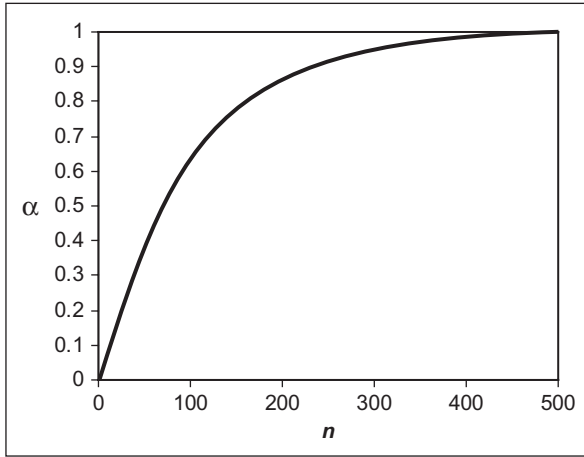


Figure 1.2 The size, α , of a surveillance system which is pursued for n time units, when the probability of a false alarm is 1% at each time point.

1.4.2 Measures of the false alarm rate

The false alarm tendency is more complicated to control in surveillance than in hypothesis testing, as was seen above (for example in Figure 1.2). There are special measures of the false alarm properties which are suitable for surveillance. The most commonly used measure is the Average Run Length when there is no change in the system under surveillance, $ARL^0 = E(t_A|D)$. A variant of the ARL is the Median Run Length, MRL.

A measure commonly used in theoretical work is the false alarm probability, $PFA = P(t_A < \tau)$. This is the probability that the alarm occurs before the change.

1.4.3 Delay of the alarm

The delay time of the detection of a change should be as short as possible. The most commonly used measure of the delay is the average run length until the detection of a true change (that occurred at the same time as the surveillance started), which is denoted by ARL^1 . The part of the definition within parentheses is seldom spelled out but generally used in the literature (see, for example, Page 1954 and Ryan 2000). Instead of the average, Gan (1993) advocates that the median run length should be used on the grounds that it may be more easily interpreted. However, also here only a change occurring at the same time as the surveillance started is considered.

In most practical situations it is important to minimize the expected delay of detection whenever the change occurs. Shiryaev (1963) suggested measures of the expected value of the delay. The expected delay from the time of change, $\tau = t$, to the time of alarm, t_A , is denoted by

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

Note that $ARL^1 = ED(1) + 1$. The $ED(t)$ will typically tend to zero as t increases. Thus, it is easier to evaluate the conditional expected delay

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

$CED(\tau)$ is the expected delay for a specific change point τ . The expected delay is generally not the same for early changes as for late ones. For most methods, the CED will converge to a constant value. This value is sometimes named the ‘steady state average delay time’ or SADT. It is, in a sense, the opposite to ARL^1 since only a very large value of τ is considered. SADT has been advocated for example by Srivastava and Wu (1993), Srivastava (1994) and Knoth (2006).

For some situations and methods the properties are about the same regardless of when the change occurs. However, this is not always true, as illustrated by Frisé and Wessman (1999). Then, it is important to consider more and other cases than just $\tau = 1$. The values of CED can be summarized in different ways. One is the maximal value over τ . Another approach is to regard τ as a random variable with the probabilities $\pi(t) = P(\tau = t)$. These probabilities can also be regarded as priors. The intensity of a change is defined as $\nu(t) = P(\tau = t | \tau \geq t)$, which is usually assumed to be constant over time. Shiryaev (1963) suggested a summarized measure of the expected delay

$$ED = E[ED(\tau)].$$

Sometimes the time available for action is limited. The Probability of Successful Detection suggested by Frisé (1992) measures the probability of detection with a delay time no longer than d

$$PSD(d, t) = P(t_A - \tau < d | t_A \geq \tau = t).$$

This measure is a function of both the time of the change and the length of the interval in which the detection is defined as successful. Also when there is no absolute limit to the detection time it is often useful to describe the ability to detect the change within a certain time. In such cases it may be useful to calculate the PSD for different time limits d . This has been done by Marshall, Best, Bottle and Aylin (2004). The ability to make a very quick detection

(small d) is important in surveillance of sudden major changes, while the long-term detection ability (large d) is more important for ongoing surveillance where smaller changes are expected.

1.4.4 Predictive value

When an alarm is called, one needs to know whether to act as if the change is certain or just plausible. To obtain this, both the risk of false alarms and the risk of delay must be considered. If τ is regarded as a random variable this can be done by one summarizing measure. The probability that a change has occurred when the surveillance method signals was suggested by Frisén (1992) as a time-dependent predictive value

$$PV(t) = P(\tau \leq t_A | t_A = t).$$

When there is an alarm ($t_A = t$), PV indicates whether there is a large probability or not that the change has occurred ($t_A \geq \tau$). Some methods have a constant PV. Others have a low PV at early alarms but a higher one later. In such cases, the early alarms will not prompt the same serious action as later ones.

1.4.5 Optimality

1.4.5.1 Minimal expected delay

Shiryayev (1963) suggested a highly general utility function, in which the expected delay of an alarm plays an important role. Shiryayev treated the case where the gain of an alarm is a linear function of the value of the delay, $t_A - \tau$, and the intensity of the change is constant. The loss associated with a false alarm is a function of the same difference. This utility can be expressed as $U = E\{u(\tau, t_A)\}$, where

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

The function $h(t_A - \tau)$ is usually a constant (say, b), since the false alarm causes the same cost of alerts and investigations irrespectively of how early the false alarm is given. In this case, we have

$$U = b P(t_A < \tau) + a_1 ED + a_2.$$

We would have a maximal utility if there is a minimal (a_1 is typically negative) expected delay from the change point for a fixed probability of a false alarm (see Section 4.3). This is termed the ED criterion. Variants of the utility function leading to different optimal weighting of the observations are suggested for example by Poor (1998) and Beibel (2000).

1.4.5.2 Minimax optimality

The minimum of the maximal expected delay after a change considers several possible change times, just like the ED criterion. However, instead of an expected value, which requires a distribution of the time of change, the least favourable value of $CED(t)$ is used. This criterion is used in Chapter 5.

Moustakides (1986) uses an even more pessimistic criterion, the ‘worst possible case’, by using not only the least favourable value of the change time, but also the least favourable outcome of $X_{\tau-1}$ before the change occurs. This criterion is very pessimistic. The CUSUM method, described in Section 1.5.3, provides a solution to the criterion proposed by Moustakides. The merits of the studies of this criterion have been thoroughly discussed for example by Yashchin (1993) and Lai (1995). Much theoretical research is based on this criterion.

1.4.5.3 ARL optimality

Optimality is often stated as a minimal ARL^1 for a fixed ARL^0 . ARL^1 is the expected value under the assumption that all observations belong to the ‘out-of-control’ distribution, whereas ARL^0 is the expected value given that all observations belong to the ‘in-control’ distribution. Efficient methods for surveillance (see Section 1.5) will put most weight on the most recent observations. Statistical inference with the aim of discriminating between the two alternatives that all observations come from either of the two specified distributions should, by the ancillarity principle, put the same weight on all observations. To use efficient methods and evaluate them by the ARL criterion is thus in conflict with this inference principle.

Pollak and Siegmund (1985) argue that for many methods, the maximal value of $CED(t)$ is equal to $CED(1)$, and with a minimax perspective this can be an argument for using ARL^1 since $CED(1) = ARL^1 - 1$. However, this argument is not relevant for all methods. In particular, it is demonstrated by Frisé and Sonesson (2006) that the maximal CED-value is not $CED(1)$ for the EWMA method in Section 1.5.5. In the case of this method, there is no similarity between the optimal parameter values according to the ARL criterion and the minimax criterion, while the optimal parameter values by the criterion of expected delay and the minimax criterion agree well.

The dominating position of the ARL criterion was questioned by Frisé (2003), since methods useless in practice are ARL optimal. The ARL can be used as a descriptive measure and gives a rough impression, but it is questionable as a formal optimality criterion.

1.4.6 Comments on evaluation measures

Computer illustrations of the interpretation of some of the measures mentioned below are made by Frisé and Gottlow (2003). Formulas for numerical approximations of some of the measures are available in literature.

1.5 General methods for aggregating information

In surveillance it is important to aggregate the available information in order to benefit from all information. This aggregation can be carried out in accordance with some general inference principles. Specific methods are then derived from the general ones. Different principles of aggregation have different properties and are thus suitable for different problems.

Some methods are highly flexible and have several parameters. The parameters can be chosen to make the method optimal for the specific conditions of the application (for example, the size of the change or the intensity of changes). Many methods for surveillance are based, in one way or another, on likelihood ratios. Thus, we will start by describing the likelihood ratio component. The likelihood ratio for a fixed value of τ is

$$L(s, t) = f_{X_s}(X_s|\tau = t)/f_{X_s}(X_s|D).$$

Most commonly used methods can be described as different combinations of these components.

1.5.1 The Shiryaev–Roberts method

The simplest way to aggregate the likelihood components is just to add them. This means that all possible times for the change up to the decision time s are given equal weight. Shiryaev (1963) and Roberts (1966) suggested the method, now called the Shiryaev–Roberts method, in which an alarm is triggered at the first time s , so that

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

where G is a constant alarm limit. This method can also be given a natural interpretation if the time of the change τ is regarded as a random variable. This method can in that case be regarded as a special case of the full likelihood ratio method. This will be further discussed in Section 1.5.6.

1.5.2 The Shewhart method

The Shewhart method (Shewhart 1931; Ryan 2000) is simple and certainly the most commonly used method for surveillance. It can be regarded as performing repeated significance tests. An alarm is triggered as soon as an observation deviates too much from the target. Thus, only the last observation is considered in the Shewhart method. An alarm is triggered at

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

where L is a constant. The alarm criterion for independent observations can be expressed by the condition $L(s, s) > G$ where G is a constant. The alarm statistic of the LR method reduces to that of the Shewhart method when $C(s) = \{\tau = s\}$ and $D(s) = \{\tau > s\}$. This is the case when we want to discriminate between a change at the current time point and the case that no change has happened yet. In this situation, we are only interested to see whether something has happened ‘now’ or not. Thus, the Shewhart method has optimal error probabilities for these alternatives for each decision time s . For large shifts, the LR method of Section 1.5.6 and the CUSUM method of Section 1.5.3 converge to the Shewhart method (Frisén and Wessman 1999 and Chapter 4). By several criteria, the Shewhart method performs poorly for small and moderate shifts. By the minimax criterion, however, it works nearly as well as the LR method for some situations.

1.5.3 The CUSUM method

The CUSUM method, first suggested by Page (1954), is closely related to the minimax criterion. Yashchin (1993), Siegmund and Venkatraman (1995) and Hawkins and Olwell (1998) give reviews of the CUSUM method. The alarm condition of the method can be expressed by the partial likelihood ratios as

$$t_A = \min\{s; \max(L(s, t); t = 1, 2, \dots, s) > G\}$$

where G is a constant. The method is sometimes called *the* likelihood ratio method, but this combination of likelihood ratios should not be confused with the full likelihood ratio method, LR.

The most commonly described application of the CUSUM method concerns the case of independent normally distributed variables. In this case, the CUSUM statistic reduces to a function of the cumulative sums

$$C_r = \sum_{t=1}^r (X(t) - \mu(t)).$$

There is an alarm for the first time s for which

$$C_s - C_{s-i} > h + ki \text{ for some } i = 1, 2, \dots, s,$$

where $C_0 = 0$ and h and k are chosen constants. In the case of a step change, the value of the parameter k is usually $k = (\mu^0 + \mu^1)/2$.

Closely related to the CUSUM method are the Generalized Likelihood Ratio (GLR) and Mixture Likelihood Ratio (MLR) methods. For the MLR method suggested by Pollak and Siegmund (1975), a prior for the shift size is used in the CUSUM method. For the GLR method, the alarm statistic is formed by maximizing over possible values of the shift (besides the maximum over possible times of the shift). Lai (1998) describes both GLR and MLR and proves a minimax result for a variant of GRL suitable for autocorrelated data.

The CUSUM method satisfies the minimax criterion of optimality described in Section 1.4.5.2. Other good qualities of the method have been confirmed for example by Srivastava and Wu (1993) and Frisén and Wessman (1999). With respect to the expected delay, the CUSUM method works almost as well as the LR and Shiryaev–Roberts methods.

1.5.4 Moving average and window-based methods

The Moving average method can be expressed by the likelihood ratios as

$$L(s, s - d) > G$$

where G is a constant and d is a fixed window width. In the standard case of normally distributed variables this will be a moving average. It will have the optimal error probabilities of the LR method when we want to detect a change which occurred at time $s - d$ (i.e. for $C = \{\tau = s - d\}$) and will thus have optimal detection abilities for changes which occurred d time points earlier.

Sometimes, as in Lai (1998), advanced methods such as the GLR method are combined with a window technique in order to ease the computational burden.

1.5.5 Exponentially weighted moving average methods

The EWMA method is a variant of a moving average method which does utilize all information. The alarm statistic is based on exponentially weighted moving averages,

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

where $0 < \lambda < 1$ and Z_0 is the target value, which is normalized to zero. The EWMA statistic gives the largest weight to the most recent observation and

geometrically decreasing weights to all previous ones. If λ is near zero, all observations have approximately the same weight. Note that if $\lambda = 1$ is used, the EWMA method reduces to the Shewhart method. The asymptotic variant, EWMAa, will give an alarm at

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

where L is a constant. In another variant of the method, EWMAe, the exact standard deviation (which is increasing with s) is used instead of the asymptotic one in the alarm limit. Sonesson (2003) found that the EWMAa version is preferable for most cases.

The EWMA method was described by Roberts (1959). Positive reports of the quality of the method are given for example by Crowder (1989), Lucas and Saccucci (1990), Domangue and Patch (1991) and Knoth and Schmid (2002). The choice of λ is important, and the search for the optimal value of λ has been of great interest in the literature. Small values of λ result in a good ability to detect early changes while larger values are necessary for changes that occur later.

Most reports on optimal values of the parameter λ refer to the ARL criterion. Friséen (2003) demonstrated that by this criterion, λ should approach zero. Methods which allocate the power to the first time points will have good ARL properties but less ability to detect a change that happens later. In fact, and wisely enough, no one seems to have suggested that λ should be chosen to zero, even though this should fulfill the ARL criterion.

The EWMA method can be seen as a linear approximation of the full LR method (see Section 1.5.6). When a change from $N(0, \sigma)$ to $N(\mu, \sigma)$ occurs with the intensity ν , the parameter λ that gives the optimality properties of the full LR method is

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

This was shown by Friséen (2003) and confirmed by large-scale simulation studies by Friséen and Sonesson (2006).

1.5.6 The full likelihood ratio method

When the time of the shift is regarded as a random variable, we can utilize this property. The full likelihood ratio method (LR) is optimal with respect to the criterion of minimal expected delay and also to a wider class of utility functions (Friséen and de Maré 1991). The full likelihood is a weighted sum of the partial likelihoods

$$L(s, t) = f_{X_s}(X_s|\tau = t)/f_{X_s}(X_s|D(s)).$$

The alarm set consists of those values of X for which the full likelihood ratio exceeds a limit. The following notation can be used: At decision time s we want to discriminate between the event $C(s) = \{\tau < s\}$ and the event $D(s) = \{\tau > s\}$. The time of an alarm for the LR method is

$$t_A = \min \left\{ s; \frac{f_{X_s}(x_s|C(s))}{f_{X_s}(x_s|D(s))} > \frac{P(\tau > s)}{P(\tau \leq s)} \cdot \frac{K}{1 - K} \right\}$$

$$= \min \left\{ s; \sum_{t=1}^s w(s, t) \cdot L(s, t) > G(s) \right\}$$

where K is a constant and $G(s)$ is the alarm limit. The time of an alarm can equivalently be written as the first time the posterior probability of a change into state C exceeds a fixed level:

$$t_A = \min\{s; P(C(s)|X_s = x_s) > K\}.$$

The posterior probability of a change has been suggested as an alarm criterion for example by Smith and West (1983). When there are only two states, C and D , this criterion leads to the LR method (Frisén and de Maré 1991). In cases where several changes may follow after each other, the process may be characterised as a hidden Markov chain and the posterior probability for a certain state may be determined (for example Harrison and Stevens 1976 and Hamilton 1989). Sometimes the use of the posterior distribution, or equivalently the likelihood ratio, is named ‘the Bayes method’. However, it depends on the situation whether the distribution of τ should be considered as a ‘prior’, as an observed frequency-distribution or if it just reflects the situation for which optimality is desired.

When the intensity, ν , of a change tends to zero, the weights $w(s, t)$ of the partial likelihoods do not depend on t , and the limit $G(s)$ of the LR method does not depend on s . Shiryayev (1963) and Roberts (1966) suggested the Shiryayev–Roberts method (mentioned in Section 1.5.1), for which an alarm is triggered at the first time s , such that

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

where G is a constant. The method can be seen as the limit of the LR method when ν tends to zero. The Shiryayev–Roberts method can also be derived as the LR method with a noninformative prior for the distribution of τ . Both the LR method and the Shiryayev–Roberts method can be expressed recursively. One valuable property of these methods is an approximately constant predictive

value Frisé and Wessman (1999), which allows the same interpretation of early and late alarms.

The LR method is optimized for the values of the change size and for the change intensity. In the case of a normal distribution, the LR method gives an alarm at

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

where the constant K determines the false alarm probability.

As mentioned before, several methods can be described by approximations or combinations of likelihood ratios (Frisé 2003). Linear approximations of the LR method are of interest for two reasons – first, for obtaining a method which is easier to use and analyse but whose properties are as good as those of the LR method, and second, for getting a tool for the analysis of the approximate optimality of other methods as in Frisé (2003).

1.6 Special aspects of surveillance for financial decisions

1.6.1 General approaches which can be used in complex situations

Situations in finance are often complex. Thus, some general approaches for surveillance in more complicated situations than those of the earlier sections are of interest. When the models are completely specified both before and after the change, the likelihood components $L(s, t)$ can usually be derived or approximated. Then, these components can be combined by any of the general information aggregation methods mentioned in Section 1.5. Lai (1995, 1998) and Lai and Shan (1999) argue that the good minimax properties of generalisations of the CUSUM method make the CUSUM suitable for complicated problems. The likelihood ratio method, LR, with its good optimality properties can also be used. Pollak and Siegmund (1985) argue that the martingale property (for continuous time) of the Shiryaev–Roberts method makes this more suitable for complicated problems than the CUSUM method. The LR method also has this property, but the CUSUM method does not.

1.6.2 Evaluation by return

The return from buying an asset at $t = 0$ and selling at time t is $r(t) = x(t) - x(0)$, where X is a monotonic function of the price.

The expected return $E[r(t_A)]$ of selling at the alarm time t_A is maximal when $E[X(t_A)]$ is maximal. Thus, a sell signal should ideally come at time τ which corresponds to a peak of the price. In Chapter 3 on technical analysis this surveillance problem is analysed.

1.6.3 Surveillance of dependent data

Financial time series often have complicated time dependencies. The theory for surveillance of dependent data is not simple. The general approaches in Section 1.6.1 can be applied to obtain methods with known optimality properties. This was made for example by Petzold *et al.* (2004). The most common approach to surveillance in the case of models with dependencies is to monitor the process of residuals. Pettersson (1998) demonstrated that for an autoregressive process, this is an approximation of the LR method. Another common approach (also used by Pettersson 1998) is to adjust the alarm limit in order to adjust the false alarm risk resulting from ignoring the dependency. Chapters 5–7 and the references in these chapters contain important contributions to this very sparsely discussed area.

1.6.4 Surveillance of discrete distributions

Most of the theory of surveillance is derived for normal distributions, but a bibliography of surveillance of attribute data is given by Woodall (1997).

1.6.5 Gradual changes

Most of the literature on surveillance treats the case of an abrupt change. In many cases in finance, however, the change is gradual. The change is thus more complicated than the standard situation of a sudden shift in a parameter from one value to another. Important characteristics should be captured by the statistic under surveillance. In the presence of a nuisance parameter, a general approach is to use a pivot statistic. Krieger, Pollak and Yakir (2003) suggest the CUSUM and Shiryaev–Roberts methods based on a statistic which does not depend on the unknown parameters in a case of an unknown prechange regression. Arteaga and Ledolter (1997) compare several procedures with respect to ARL properties. One of the suggestions in the paper is a window method based on the likelihood ratio and isotonic regression techniques. In general, window methods (see Section 1.5.4) are inefficient for detecting gradual changes (Järpe

2000). Yashchin (1993) discusses generalizations of the CUSUM and EWMA methods to detect both sudden and gradual changes.

It may be hard to model the shape of a gradual change exactly or even to estimate the baseline accurately. Then, the timely detection of a change in monotonicity is of interest. The start of an increase is of course of special interest, but also the decline may be of interest to get timely sell and buy signals.

When the knowledge on the shape of the curve is uncertain, nonparametric methods are of interest. Frisén (2000) suggested surveillance that is not based on any parametric model but only on monotonicity restrictions. This surveillance method was described and evaluated by Andersson (2002) and is further described in Chapter 3.

1.6.6 Changes between unknown levels

After a change, the level of the statistic under surveillance (for example the variance) is seldom known. However, this is not a serious problem. The false alarm properties will remain the same even if the level after the change is not known. The method could be designed to be optimal for a change of a specific size, but this is not required. The unknown parameters can be handled within different frameworks corresponding to different restrictions on possible optimality.

To control false alarms is usually more important than to optimize the detection ability. Knowledge of the prechange conditions is important. The baseline is often estimated and used as a plug-in value in the method. The estimated baseline value will affect the performance of the method. In the situation where we want to detect an increase, we will get more false alarms if the baseline is underestimated than if the true value had been used. The opposite is true if the baseline is overestimated.

One way to avoid the problem of unknown parameters is to transform the data to invariant statistics. Frisén (1992) and Sullivan and Jones (2002) use the deviation of each observation from the average of all previous ones. Gordon and Pollak (1997) use invariant statistics combined by the Shiryaev–Roberts method to handle the case of an unknown prechange mean of a normal distribution. Krieger *et al.* (2003) use invariant statistics combined by the CUSUM and Shiryaev–Roberts methods for surveillance of a change in regression.

When both the baseline and the change are unknown, the aim of the surveillance could be to detect a change in a stochastically larger distribution. Bell, Gordon and Pollak (1994) suggested a nonparametric method geared to the exponential distribution. The nonparametric method of Chapter 3, designed for

the detection of a change in monotonicity, also avoids the problem of unknown values of the baseline and the change.

The use of the maximum difference (measured for example by the likelihood ratio) between the baseline and the changed level is a useful approach. The GLR method (Lai 1995, 1998) uses the maximum likelihood estimator of the value after the change. Kulldorff (2001) used the same technique for the detection of clustering in spatial patterns.

Another general approach for unknown levels is the Bayesian one. The MLR method suggested by Pollak and Siegmund (1975) uses priors for the unknown parameters in the CUSUM method. Lawson (2004) used priors for the unknown parameters to calculate the posterior means in a Bayesian space-time interaction model.

1.6.7 Multivariate surveillance

We may have several data streams containing information. This is the case for example in portfolio optimization. We may also have several statistics, such as both the mean and the variance, to monitor. In Chapters 5–7 multivariate techniques for financial problems are extensively discussed.

If the model can be completely specified both before and after the change, then it is possible to derive the likelihood components $L(s, t)$ and aggregate them by a method which guarantees optimality. In complicated problems, however, this is seldom realistic. Instead, a reduction of the multivariate surveillance problem is common (Sonesson and Frisén 2005).

A reduction of the dimensionality of the problem is a natural first approach. Principal components could be used to reduce the dimensionality, but Lowry and Montgomery (1995) argued that unless the principal components can be interpreted, a surveillance method based on them may be difficult to interpret. In Rosolowski and Schmid (2003) and Chapters 5–7 the Mahalanobis distance is used to reduce the dimensionality of the statistic, thus expressing the distance from the target of the mean and the autocorrelation in a multivariate time series.

The most common way to handle multivariate surveillance is to reduce the information to one statistic and then monitor this statistic in time. Wessman (1998) proved that this is a sufficient reduction when changes occur simultaneously in all variables.

Another commonly used approach is to make parallel surveillance for each variable and make a general alarm when there is an alarm for any of the components, see Stoumbos, Reynolds Jr, Ryan and Woodall (2000). Any univariate surveillance method could be used. Parallel CUSUM methods were used by Marshall *et al.* (2004). The false alarms were controlled by using the False Discover Rate (FDR) from Benjamini and Hochberg (1995). For evaluating

the detection ability, the probability of successful detection (see Section 1.4.3) was used.

The more advanced approach of vector accumulation is an in-between of the reduction by time and the reduction by variable. Here the accumulated information on each component is used to transform the vector of component-wise alarm statistics into a scalar alarm statistic and make an alarm if this statistic exceeds a limit, see for example Rogerson and Yamada (2004). It is also possible to construct the multivariate method while aiming at satisfying some global optimality criterion. Järpe (1999) suggested an ED optimal surveillance method of clustering in a spatial log-linear model. Lowry, Woodall, Champ and Rigdon (1992) proposed a multivariate extension of the univariate EWMA method, which is referred to as MEWMA. This can be described as the Hotelling T^2 control chart applied to univariate EWMA statistics instead of the original data from only the current time point and is thus a vector accumulation method. Crosier (1988) suggested the MCUSUM method, where a statistic consisting of univariate CUSUMs for each component is used. This is similar to the MEWMA statistic, which corresponds to a vector accumulation method. However, the way in which the components are used is not the same. An alternative way to construct a vector accumulating multivariate CUSUM is given by Pignatiello and Runger (1990). The methods use different weighting of the variables. One important feature of these two methods is that the characteristic zero-return of the CUSUM technique is constructed in a way that is suitable when all the components change at the same time point.

Different aspects on approaches for multivariate surveillance were given by Frisé (2003) and Sonesson and Frisé (2005). The multivariate methods can be evaluated by the measures and criteria described above or by generalized measures. Wessman (1999) suggested a generalization of the ARL measure to allow for the possibility of different change times for different variables. Controlling the false discovery rate is of interest when making conclusions about several variables and is used for example by Wong, Moore, Cooper and Wagner (2003). However, the question of optimality is always complex in multidimensional cases.

1.7 Content and the relation between chapters

Each chapter is self-contained. It is thus not necessary to read the chapters in the order they appear in the book. In this section, the relation between the chapters will be described in order to facilitate the choice of reading order. This section will also demonstrate the wide spectrum of problems and methods

in financial surveillance. The short description of each chapter given here may also give information on which chapter to choose for a specific purpose.

Different financial series are analysed in the different chapters. In Chapter 3 the Hang Seng index is analysed with respect to optimal times for trading. This is a market-value weighted index of the 33 largest companies on the Hong Kong stock market. In Chapter 4 a period of Standard and Poor's 500 stock market index is analysed to investigate the ability of some suggested methods to detect a documented change in volatility. In Chapters 5 and 6 there are several analyses of the Morgan Stanley Capital International index (MSCI) for Germany. In Chapter 5 also the MSCI indices for the US and Japan are analysed. In Chapter 7 the optimal allocations of a portfolio of the US market, a non-US index, oil and gold are determined. In addition, most chapters contain simulation studies to further describe the properties of the suggested methods.

In finance much effort has been made to find models which are suitable for financial time series. Special demands have to be met. Chapter 2 gives an overview of problems in finance and models which have been found useful to describe financial data for various purposes. However, Chapter 2 is somewhat more mathematically demanding than the other chapters. Thus, it might be a good idea to start by just taking a quick look at Chapter 2, as an introduction to the kind of models which are considered to capture the characteristics of financial data. A more thorough reading of Chapter 2 can be made after the other chapters. It will then serve as an overview of a great number of models. This gives the possibility to compare methods. Although this book does not discuss surveillance with reference to all these models, it might inspire new research on the development of methods for surveillance of financial models of great interest. This will be discussed in Chapter 9. Some of the models described in Chapter 2 will be used later. However, chapters where a specific model is used will contain a description of that model. Chapters 3–8 each give a description of surveillance with reference to some specific models. Each chapter on surveillance contains all the necessary information about the model used in that chapter. However, it might be a good idea to look also at the part of Chapter 2 where the specific model and similar ones are treated.

Chapter 3 links traditional technical analysis to the new methods for surveillance. Even though the chapter contains new research material it is not very demanding mathematically. Methods for traditional technical analysis could be ad hoc without derivation from financial or statistical science. Thus, methods for technical analysis are not always considered scientific. New mathematical methods have a better reputation even though they are sometimes based on assumptions which might be doubted. Technical analysis has a background in

common sense about how to find the right time to make transactions in order to increase profit. In Chapter 3 these methods are justified by a demonstration of how some common methods can be derived from simple stochastic models, natural expression of a good return and good methods for surveillance. Thus, common methods for technical analysis are linked to surveillance theory. In this way one can also find optimal ways to maximize the return for the simple models as well as ways to improve the methods of technical analysis by using a bit more complicated methods. This will be further discussed in Chapter 9.

In technical analysis the history of the price of a stock is assumed to contain information, and technical analyses are methods for extracting this information. The natural aim of technical analysis is to maximize the return. The possibility of achieving this has been much doubted during the last decades. The theory of stochastic finance has been based on an assumption of an efficient market where the financial markets are arbitrage-free and there is no point in trying to determine the optimal transaction time. Even though this view is generally accepted nowadays there are also some doubts that it is generally applicable. When the information about the process is incomplete, as for example when a change point could occur, there may be an arbitrage opportunity, as demonstrated by Shiryaev (2002). Also, there are agents who reject the efficient market hypothesis, and some studies give support for the profitability of the prospective framework. Technical analysis is used in practice, and an analysis of its relation to other models and methods may thus be of interest.

When the efficient market hypothesis is accepted, the interest is drawn to the variability of the price. A low variability is desired. Methods for trading strategies which minimize the variance in different situations are described in Chapters 4–7. In Chapter 4 an iid process (except for a step change in the variance) is assumed, and a sufficient statistic for the ordinary variance is under surveillance. When time dependence is allowed, the measure of variance becomes more complicated. In Chapter 5 the variance in a linear time series is under surveillance. In Chapter 6 nonlinear time series (GARCH, generalized autoregressive conditional heteroscedasticity) models are studied. In Chapter 7 surveillance of the minimum variance portfolio is discussed.

In the theory of surveillance, detection of a parametric step change in the model is the most developed part. While the other chapters deal with this problem, Chapter 3 deals with a nonparametric approach to detect a gradual change.

Most methods for surveillance are based on independent observation (given τ). In Chapters 5–7 several general techniques to obtain good methods for dependent time series are used. In Chapter 5 the impact of neglecting the time series structure is examined within an extensive simulation study.

In Chapters 3 and 4, the methods are expressed by likelihood ratios and the time of change, τ , is regarded as a stochastic variable. This facilitates the construction of methods which are optimal with respect to the expected delay (ED), as described in Section 1.2 of this chapter. This gives a summarizing measure of timeliness. In the other chapters the change point is regarded as a fix but unknown constant. In these chapters evaluations are made for different values of τ and especially for $\tau = 1$ which gives the commonly used measure ARL^1 . This approach avoids assumptions on the distribution of τ . The delay for the least favourable value of τ corresponds to the minimax criteria for which the CUSUM method has optimality properties.

Even though the statistics studied may be quite different, all methods use some way of aggregating the information over time in order to enable a timely decision. The methods differ as to the way in which the different partial likelihood ratios are weighted, as described in Section 1.2 of this chapter. In Chapter 3 a new maximum likelihood method is compared to the hidden Markov method, HMM, and the CUSUM method. In Chapter 4 the methods studied are the full likelihood ratio, Shiryaev–Roberts, Shewhart and CUSUM methods. Chapters 5–7 are dominated by the EWMA approach, but also the CUSUM method is evaluated in Chapter 5. The EWMA statistic cannot be expressed directly, but Friséen (2003) demonstrated that it can be seen as an approximation of the likelihood ratio method.

In Chapters 5–7 the problem of multivariate surveillance is treated while the other chapters deal with univariate surveillance. Several general techniques for multivariate surveillance are adopted in order to derive appropriate methods. The trading strategy may also involve the optimal times for updating the portfolio. The portfolio problem is treated in Chapter 7, where an optimal method and a case study of the multivariate problem of portfolio optimization are given.

Chapter 8 deals with surveillance of a continuous time model in contrast to the other chapters, which analyse discrete time models. Continuous time models are popular for describing some financial processes. However, since the models tend to be complicated, there is still a lack of methods for surveillance of continuous financial series. The estimation of parameters which can serve as indicators of important changes may be hard to achieve. Here, computer programs are provided for this task. Even though the model is specified in continuous time, the decisions are, in practice, made in discrete time. The influence of the time between observations and other problems of continuous time surveillance are discussed in Chapter 8. Decisions in continuous time are in fact easier than decisions in discrete time, since the latter type causes an overshoot. Many results are first derived for continuous time and only later for

the discrete time which applies in practice (see for example Shiryaev 1963 and Srivastava and Wu 1993).

In Chapter 9 some conclusions are given and the prospect of future progress is discussed. Some open problems for further research are suggested. The level as regards mathematics and citations to scientific literature is designed to be a reasonable starting point for statisticians and quantitatively minded people in finance. Knowledge of statistical concepts corresponding to about one year of study is required for understanding the text.

A companion website to this book is available at: <http://www.wiley.com/go/frisen>

References

- Andersson, E. (2002). Monitoring cyclical processes – a nonparametric Approach. *Journal of Applied Statistics*, **29**, 973–990.
- Arteaga, C. and Ledolter, J. (1997). Control charts based on order-restricted tests. *Statistics & Probability Letters*, **32**, 1–10.
- Beibel, M. (2000). A note on sequential detection with exponential penalty for the delay. *The Annals of Statistics*, **28**, 1696–1701.
- Bell, C., Gordon, L. and Pollak, M. (1994). An efficient nonparametric detection scheme and its application to surveillance of a Bernoulli process with unknown baseline. In *Change-Point Problems*, eds. E. Carlstein, H.-G. Muller and D. Siegmund, Hayward, California: IMS Lecture Notes, Monograph Series, pp. 7–27.
- Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society B*, **57**, 289–300.
- Bock, D. (2008). Aspects on the control of false alarms in statistical surveillance and the impact on the return of financial decision systems. *Journal of Applied Statistics*, **35**.
- Chu, C.-S. J., Stinchcombe, M. and White, H. (1996). Monitoring structural change. *Econometrica*, **64**, 1045–1065.
- Cizek, P., Härdle, W. and Weron, R. (eds) (2005). *Statistical Tool for Finance and Insurance*. Springer-Verlag, Berlin.
- Crosier, R. B. (1988). Multivariate generalizations of cumulative sum quality-control schemes. *Technometrics*, **30**, 291–303.
- Crowder, S. V. (1989). Design of exponentially weighted moving average schemes. *Journal of Quality Technology*, **21**, 155–162.
- Domangue, R. and Patch, S. C. (1991). Some omnibus exponentially weighted moving average statistical process monitoring schemes. *Technometrics*, **33**, 299–313.
- Franke, J., Härdle, W. and Hafner, C. (2004). *Statistics of Financial Markets. An Introduction*. Springer-Verlag, Berlin.
- Frisén, M. (1992). Evaluations of methods for statistical surveillance. *Statistics in Medicine*, **11**, 1489–1502.

- Frisén, M. (1994). Statistical surveillance of business cycles, Technical Report 1994:1 (Revised 2000).
- Frisén, M. (2003). Statistical surveillance. Optimality and methods. *International Statistical Review*, **71**, 403–434.
- Frisén, M. and de Maré, J. (1991). Optimal Surveillance. *Biometrika*, **78**, 271–280.
- Frisén, M. and Gottlow, M. (2003). Graphical evaluation of statistical surveillance. Technical Report Research Report 2003:10, Statistical Research Unit, Göteborg University.
- Frisén, M. and Sonesson, C. (2006). Optimal surveillance based on exponentially weighted moving averages. *Sequential Analysis*, **25**, 379–403.
- Frisén, M. and Wessman, P. (1999). Evaluations of likelihood ratio methods for surveillance. Differences and Robustness. *Communications in Statistics, Simulation and Computation*, **28**, 597–622.
- Föllmer, H. and Schied, A. (2002). *Stochastic Finance. An Introduction in Discrete Time*. de Gruyter, Berlin.
- 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.
- Gordon, L. and Pollak, M. (1997). Average run length to false alarm for surveillance schemes designed with partially specified pre-change distribution. *The Annals of Statistics*, **25**, 1284–1310.
- Gourieroux, C. and Jasiak, J. (2002). *Financial Econometrics: Problems, Models and Methods*. University Presses Of California, Columbia and Princeton, New Jersey.
- Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, **57**, 357–384.
- Harrison, P. J. and Stevens, C. F. (1976). Bayesian Forecasting, with discussion. *Journal of the Royal Statistical Society B*, **38**, 205–247.
- Hawkins, D. M. and Olwell, D. H. (1998). *Cumulative Sum Charts and Charting for Quality Improvement*. Springer-Verlag, New York.
- Härdle, W., Kleinow, T. and Stahl, G. (eds) (2002). *Applied Quantitative Finance. Theory and Computational Tools*. Springer-Verlag, New York.
- Järpe, E. (1999). Surveillance of the interaction parameter in the Ising model. *Communications in Statistics. Theory and Methods*, **28**, 3009–3025.
- Järpe, E. (2000). On univariate and spatial surveillance. PhD thesis. Göteborg University, Department of Statistics.
- Knuth, S. (ed.) (2006). *The Art of Evaluating Monitoring Schemes – How to Measure the Performance of Control Charts?* (Vol. 8), eds. H.-J. Lenz and P.-T. Wilrich, Physica Verlag, Heidelberg.
- Knuth, S. and Schmid, W. (2002). Monitoring the mean and the variance of a stationary process. *Statistica Neerlandica*, **56**, 77–100.
- Krieger, A. M., Pollak, M. and Yakir, B. (2003). Surveillance of a simple linear regression. *Journal of the American Statistical Association*, **98**, 456–469.
- Kulldorff, M. (2001). Prospective time periodic geographical disease surveillance using a scan statistic. *Journal of the Royal Statistical Society A*, **164**, 61–72.
- Lai, T. L. (1995). Sequential changepoint detection in quality-control and dynamical systems. *Journal of the Royal Statistical Society B*, **57**, 613–658.

- Lai, T. L. (1998). Information bounds and quick detection of parameters in stochastic systems. *IEEE Transactions on Information Theory*, **44**, 2917–2929.
- Lai, T. L. and Shan, Z. (1999). Efficient recursive algorithms for detection of abrupt changes in signals and control systems. *IEEE Transactions on Automatic Control*, **44**, 952–966.
- Lawson, A. (2004). Some considerations in spatial-temporal analysis of public health surveillance data. In *Monitoring the Health of Populations: Statistical Principles & Methods for Public Health Surveillance*, eds R. Brookmeyer and D. F. Stroup. Oxford University Press, Oxford, pp. 289–314.
- Lowry, C. A. and Montgomery, D. C. (1995). A review of multivariate control charts. *IIE Transactions*, **27**, 800–810.
- Lowry, C. A. Woodall, W. H., Champ, C. W. and Rigdon, S. E. (1992). A multivariate exponentially weighted moving average control chart. *Technometrics*, **34**, 46–53.
- Lucas, J. M. and Saccucci, M. S. (1990). Exponentially weighted moving average control schemes: Properties and enhancements. *Technometrics*, **32**, 1–12.
- Marshall, C., Best, N., Bottle, A. and Aylin, P. (2004). Statistical issues in the prospective monitoring of health outcomes across multiple units. *Journal of the Royal Statistical Society A*, **167**, 541–559.
- Page, E. S. (1954). Continuous inspection schemes. *Biometrika*, **41**, 100–114.
- Pettersson, M. (1998). Monitoring a freshwater fish population: Statistical surveillance of biodiversity. *Environmetrics*, **9**, 139–150.
- Petzold, M., Sonesson, C., Bergman, E. and Kieler, H. (2004). Surveillance in longitudinal models. Detection of intrauterine growth retardation. *Biometrics*, **60**, 1025–1033.
- Pignatiello, J. J. and Runger, G. C. (1990). Comparisons of multivariate CUSUM charts. *Journal of Quality Technology*, **22**, 173–186.
- Pollak, M. and Siegmund, D. (1975). Approximations to the expected sample size of certain sequential tests. *The Annals of Statistics*, **3**, 1267–1282.
- Pollak, M. and Siegmund, D. (1985). A diffusion process and its applications to detecting a change in the drift of Brownian motion. *Biometrika*, **72**, 267–280.
- Poor, V. H. (1998). Quickest detection with exponential penalty for delay. *The Annals of Statistics*, **26**, 2179–2205.
- Roberts, S. W. (1966). A comparison of some control chart procedures. *Technometrics*, **8**, 411–430.
- Rogerson, P. A. and Yamada, I. (2004). Monitoring change in spatial patterns of disease: Comparing univariate and multivariate cumulative sum approaches. *Statistics in Medicine*, **23**, 2195–2214.
- Rosolowski, M. and Schmid, W. (2003). EWMA charts for monitoring the mean and the autocovariances of stationary Gaussian processes. *Sequential Analysis*, **22**, 257–285.
- Ryan, T. P. (2000). *Statistical Methods for Quality Improvement* (2nd edn). John Wiley & Sons, Ltd, New York.
- Scherer, B. and Martin, R. D. (2005). *Introduction to Modern Portfolio Optimization with Nuopttm, S-Plus and S⁺ Bayestm*. Springer-Verlag, New York.
- Schmid, W. and Schöne, A. (1997). Some properties of the EWMA control chart in the presence of autocorrelation. *The Annals of Statistics*, **25**, 1277–1283.
- Shewhart, W. A. (1931). *Economic Control of Quality of Manufactured Product*. MacMillan and Co., London.

- Shiryayev, A. N. (1963). On optimum methods in quickest detection problems. *Theory of Probability and Its Applications*, **8**, 22–46.
- Shiryayev, A. N. (1999). *Essentials of Stochastic Finance: Facts, Models, Theory*. World Scientific, Singapore.
- Shiryayev, A. N. (2002). Quickest detection problems in the technical analysis of financial data. In *Mathematical Finance – Bachelier Congress 2000*, eds. H. Geman, D. Madan, S. Pliska and T. Vorst. Springer-Verlag, Berlin.
- Siegmund, D. and Venkatraman, E. S. (1995). Using the generalized likelihood ratio statistic for sequential detection of a change-point. *The Annals of Statistics*, **23**, 255–271.
- Smith, A. F. and West, M. (1983). Monitoring renal transplants: An application of the multiprocess Kalman filter. *Biometrics*, **39**, 867–878.
- Sonesson, C. (2003). Evaluations of some exponentially weighted moving average methods. *Journal of Applied Statistics*, **30**, 1115–1133.
- Sonesson, C. and Frisén, M. (2005). Multivariate surveillance. In *Spatial Surveillance for Public Health*, eds. A. Lawson and K. Kleinman. John Wiley & Sons, Ltd, New York. pp. 169–186.
- Srivastava, M. S. (1994). Comparison of CUSUM and EWMA procedures for detecting a shift in the mean or an increase in the variance. *Journal of Applied Statistical Science*, **1**, 445–468.
- Srivastava, M. S. and Wu, Y. (1993). Comparison of EWMA, CUSUM and Shiryayev–Roberts procedures for detecting a shift in the mean. *The Annals of Statistics*, **21**, 645–670.
- Stoumbos, Z. G., Reynolds Jr, M. R., Ryan, T. P. and Woodall, W. H. (2000). The state of statistical process control as we proceed into the 21st century. *Journal of the American Statistical Association*, **95**, 992–998.
- Sullivan, J. H. and Jones, L. A. (2002). A self-starting control chart for multivariate individual observations. *Technometrics*, **44**, 24–33.
- 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). Studies on the surveillance of univariate and multivariate processes. PhD thesis, Göteborg University, Department of Statistics.
- Wong, W. K., Moore, A., Cooper, G. and Wagner, M. (2003). WSARE: Whats Strange About Recent Events? *Journal of Urban Health*, **80**, 166–175.
- Woodall, W. H. (1997). Control charts based on attribute data: bibliography and review. *Journal of Quality Technology*, **29**, 172–183.
- Yashchin, E. (1993). Statistical control schemes – methods, applications and generalizations. *International Statistical Review*, **61**, 41–66.

