Research Report
Department of Statistics
Göteborg University
Sweden

Statistical surveillance
of business cycles

Marianne Frisén

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

STATISTICAL SURVEILLANCE OF BUSINESS CYCLES[1]

MARIANNE FRISÉN

Methods for timely detection of turning-points in business cycles are discussed from a statistical point of view. The theory on optimal surveillance is used to characterize different approaches advocated in literature. This theory is also used to derive a new method for nonparametric detection of turning-points. It utilizes the characteristics of monotonic and unimodal regression. Estimation of parameters in a more or less stable model is thus avoided. Different new ways to evaluate methods are used and discussed. The principles are illustrated by data from Sweden and the USA.

KEYWORDS: Early warning; monitoring, index of leading indicators, business cycle, turning-point, optimal, likelihood ratio, nonparametric, unimodal regression, monotonic regression

1. INTRODUCTION

Decisions about phases of business cycles are important in government as well as in industry. Government policy programs of job creation through public works or public service employment have been repeatedly called counter-cyclical without in fact being so. Most such programs came, according to Zarnowitz and Moore (1982), into effect much too late. The tardiness of policies designed to stimulate employment induces some unintended effects. Öller (1987) finds serious lack of timeliness in the Finnish policy programs. According to Westlund (1993) the skill and accuracy with which business cycle forecasting is made determine the success or failure of governmental programs as well as of management decisions.

Business cycle forecasting has been of increasing interest in recent years. For reviews and general discussions see Zarnowitz (1985), Westlund and Zackrisson (1986), Hackl and Westlund (1989), Hackl and Westlund (1991), Lahiri and Moore (1991) Westlund (1993) and Birchenhall et.al.(1999).

The techniques actually in use differ between countries. There are diverging opinions on the importance of models based on economic theories versus models based on fitting of flexible standard models (e.g. ARIMA models). While Edlund and Sogaard (1993) advocate the flexible models, Öller (1990) and Lee and Shields (2000) discuss the advantages of statistics derived from qualitative surveys. Most suggestions include a little of both and involve leading indicators.

Several papers, i.e. Neftci (1982), Stekler (1991) and Nazmi (1993) give arguments for the important difference between turning-point prediction and forecasting of the precise value of the cyclical variable. Among others, Zarnowitz and Braun in the

volume by Stock and Watson (1994) commented upon the bad fit near turning points for many forecasting methods. A statistical model, which results in impressively close fits, may fail to signal turns but compensate by providing excellent fits for observations between turning-points. The possibly asymmetric nature of the business cycle has been suggested as an explanation for this. The stochastic behaviour could be different before and after the turn. Another explanation is that the inferential problems are different and that a method that suits one kind of problem might not be optimal for a different (even though related) problem.

Turning-point forecasting of business cycles is usually based on an identification of turning-points in a leading index (e.g. the papers in the collection by Stock and Watson 1994). Also, this paper concentrates on this approach.

Here the emphasis will be on different approaches to the inferential aspects of the repeated decisions rather than on different models used. Even though there is an important difference between business growth peaks and business peaks this distinction is not made here. The paper deals with the special inferential problems caused by a successively increasing number of observations and the need of successive decisions. There is no fixed hypothesis to be tested. At time t the hypothesis that no change of phases has occurred before t is of interest. At time t+1 the main concern might be whether a change has occurred before that time (t+1).

In many different areas there is a need of continual observation of time series, with the goal of detecting an important change in the underlying process as soon as possible after it has occurred. The timeliness of decisions is taken into account in the vast literature on quality control charts. In that area it is often important with simplicity, while in the economic analyses accuracy and timeliness are more important. Also, the literature on stopping rule procedures

is relevant. However, business cycle forecasting is too complicated to be handled by one optimality criterion alone.

Here the special problems of changes of phases of business cycles will be treated by a comparison with results from other areas. Also, the vocabulary needs attention. It will for example be demonstrated that the method called CUSUM in the econometric literature is not identical with the method with that name in quality control. On the other hand, the method by Andrews (1993) can be regarded as a special case of the CUSUM method as described by Siegmund (1985) in the context of sequential analysis.

In Section 2 there is a review of some general results on statistical methods for surveillance, of relevance for the following discussion. Also the inferential frame and notations and specifications for the detection of change in business phases are given in that section. In Section 3 methods specially designed for detection of changes in business phases are treated. The use of leading indicators are discussed in Section 3.1. In Section 3.2 an overview of methods and their relation to different optimality criteria and to general classical methods are discussed. In Section 3.3 data on the Swedish economy during January 1960 to April 1987 are used to illustrate the discussion. In Section 3.4 a new robust approach is suggested. It is based on results on nonparametric regression with qualitative restrictions. In Section 4 different ways of evaluations are discussed and Swedish and American data are used for illustration. In Section 5 some concluding remarks are made.

2. STATISTICAL METHODS FOR SURVEILLANCE

A wide variety of methods have been suggested in the general literature on surveillance, see, e.g. Zacks (1983) Wetherill and Brown (1990) and Frisén and de Maré (1991). Some methods (like the Shewhart test) only take the last observation into account. Others (simple sums or averages) give the same weight to all observations. For most applications it is relevant to use something in between. That is, all observations are taken into account but more weight is put on recent observations than on old ones. The CUSUM and the EWMA are such methods. They are much discussed and both are nowadays often recommended. Both these methods include the extremes mentioned above as special cases and the relative weight on recent observations and old ones can be continuously varied by varying their two parameters.

2.1 Inferential framework

In practice, turning-points are identified after eliminating long-term trends and seasonal effects, and often after smoothing short-term irregular variations. As was mentioned in the introduction, the emphasis here will be on the inferential aspects of the repeated decisions rather than on modeling. In order to focus on the inferential matters all details which are relevant for business cycle modeling will not be analyzed. For example, here only those aspects about how the seasonal adjustment is made that influence on optimality or the sequential nature of the inference will be considered.

The observation vector at time t is named X(t). It might be a multivariate random variable consisting of, e.g. one component Y(t) of main interest (for example an index of industrial production) and another component L(t) (for example a vector of leading indicators). The observation vector might also be replaced by a

vector of sufficient statistics (one component might for example be a recursive residual from an econometric model). The data available for the decision at time s are $X_s = \{X(t): t = 1,2,...s\}$.

There is a stochastic process which determines the state of the system. We might have a special interest in identifying states which are phases of the leading index. In the phases of expansion we have that $I(t) = E[L(t) - L(t-1)] > 0$. In the phases of recession we have $I(t)<0$.

The critical event of interest at decision time s is denoted $C(s)$. The critical event might be a change of phases. If the business cycle has been in an expansion phase, the critical event might be a change, at time τ, to a recession phase. Then, $I(t) > 0$ for $t= 1,...,\tau-1$ and $I(t) < 0$ for $t= \tau, \tau+1,...$ The aim might be to discriminate between the case where the change has happened and the case where it has not happened yet. The aim, at time s, is then to discriminate between

$$C(s) = \{\tau \leq s\} = \{I(s)<0\} \quad \text{and} \quad D(s) = \{\tau>s\} = \{I(s)>0\}.$$

Sometimes only recent changes (within the time limit d) are of real interest. The aim is then to discriminate between

$$C(s) = \{s-d<\tau \leq s\} \quad \text{and} \quad D(s) = \{\tau>s\}.$$

The time point τ where the critical event occurs might be regarded as a random variable with the density $P(\tau=t) = \pi_t$. The intensity q_t of a change is

$$q_t = P(\tau=t \mid \tau \geq t)$$

For business cycles a constant transition probability has been advocated by, i.e. McCulloch (1975), Diebold and Rudebusch (1990), (1991) and Hamilton (1993). Neftci (1982) on the other

hand estimated the details of how the transition probability depends on the time since last transition.

The aim is to discriminate between the states of the system at each decision time s, s=1,2,... by the observation $X_s = \{X(s): t \leq s\}$. We will consider different ways to construct alarm sets A(s) with the property that when X_s belongs to A(s) it is an indication that C(s) occurs.

The distinction made in Frisén and de Maré (1991) between active and passive alarm is relevant for business cycles. We use an alarm set A=A(s) which is optimal to discriminate between the two events C(s) and D(s) based on the information available at time s. At time s+1 we will discriminate between the two events C(s+1) and D(s+1). At time s+1 our actions at time s may or may not have affected the distribution of the processes. In the case when our actions at an earlier time point do not affect the distributions we say that we have passive surveillance. In the case where the surveillance will be ended as soon as an alarm occurs, we call it active surveillance. In the surveillance of business cycles we might for some purposes study the properties of a method of surveillance as if our alarms did not affect the process. For other purposes we might consider only the time up to the first alarm (this is the usual approach in econometrics) and thus regard the surveillance as active.

Another distinction is that between a single decision and a sequence of decisions. At a single decision alarm, an ordinary test is natural. For a sequence of decisions characteristics of the sequence (such as expected delay to alarm) become interesting. In the econometric literature CUSUM-variants are usually used for a fixed series and evaluated by ordinary level of significance and power even when the problem might be one of sequential decisions (i.e. Brown, Durbin and Evans 1975, Ploberger and Krämer 1990 and Shukur 1993). The local power is treated by Krämer, Ploberger

and Alt (1988) and Nyblom (1990). A further discussion on CUSUM variants in economics will be given in Section 3.2.

2.2 Optimality

Different kinds of optimality criteria for surveillance were analysed by Frisén and de Maré (1991). For some of these criteria a probability distribution of the time of the change, τ, is considered and summarizing measures over this distribution are used. Other approaches and their consequences were also discussed.

The problem of finding the method that maximizes the detection probability for a fixed false alarm probability, and a fixed decision time, was treated by de Maré (1980) and Frisén and de Maré (1991). A likelihood-ratio (LR) method is the solution to this criterion. According to the LR method we have an alarm as soon as the likelihood ratio exceeds a limit, that is

$$f_{X(s)}(x(s) \mid C) / f_{X(s)}(x(s) \mid D) = p(x_s) > k_s.$$

An application of the LR method to the detection of turning-points of the business cycles will be given in Section 3.4.

Different kinds of utility functions were also discussed by Frisén and de Maré (1991). An important specification of utility is that of Girshick and Rubin (1952) and Shiryaev (1963). They treat the case where the gain of an alarm is a linear function of the difference τ-t_A between the time of the change and the time of the alarm. The loss of a false alarm is a function of the same difference. Their solution to the maximisation of the expected utility is identical to the LR method with constant limit k_s=k.

The posterior distribution PD(s) = $P(C(s) \mid X_s)$ has been suggested as an alarm statistic by, e.g. Neftci (1982), Smith et al. (1983), Hamilton (1989), Diebold and Rudebusch (1989) and Jun and Joo (1993). This statistic was demonstrated by Frisén and de

Maré (1991) to be equivalent to that of the LR method when there are only two states C and D. Here a continued recession (or expansion) and a change to expansion (or recession) would be two such states D and C.

In order to make policy programs that are counter-cyclical there is a limited time available for actions. Then, the expected value of the difference $\tau\text{-}t_A$ is not of main interest. Instead of using the expected value as above, the probability that the difference not exceeds a fixed limit is used. The fixed limit, d, is the time available for successful detection. This probability (as a function of the time of the change) was suggested by Frisén (1992) as a measure of the performance. Bojdecki (1979) considered the expected value over the time of the change and suggested as optimality criterion the maximum of

$$P(|\tau - t_A|) \leq d)$$

The moving average method (with window width d) which is often used in econometrics can be shown to be a special case of the solution of Bojdecki (1979).

A minimax criterion for the worst possible value of τ might be used. Minimax solutions to the problem of minimizing the difference $\tau\text{-}t_A$ between the time of the change and the time of the alarm avoid the requirement of information about the distribution of τ. Pollak (1985) gives an approximate solution for the case of the worst value of τ. The solution is a randomized procedure that would hardly be used in practice. For most applications however it would be more appropriate with a method that depends on τ than one that depends on an ancillary random procedure.

Moustakides (1986) used a still more pessimistic criterion by using not only the worst time τ but also the worst possible outcome before the change occurs. The CUSUM method below is (except for the first time point) the solution to the criterion posed by

Moustakides. Ritov (1990) motivates CUSUM by a minimax criterion which is similar but not identical to that of Moustakides. In economics, CUSUM-methods are frequently used and discussed. The CUSUM-method (as defined in the literature on quality control) is a minimax method and also a natural (but not optimal) combination of optimal sub-methods, as was shown by Frisén and de Maré (1991).

3. METHODS FOR DETECTION OF TURNING-POINTS IN BUSINESS CYCLES

3.1 Leading indicators

In business cycle modeling and forecasting leading indicators are often used to predict the business cycle reference series. Turning-point forecasting is usually based on an identification of turning-points in the leading index. According to Nazmi (1993) the leading indicator technique has outlived the rise and fall of the large-scale econometric modeling approach to business forecasting. In order to concentrate on the inferential matters we will mainly look at the situation where an index of industrial production Y(t) is closely related to a leading index L(t-lag) where lag is the time-lag. An identification of a turning-point in the leading index can then be used to predict that the industrial production will have a turning-point at a time lag after the change of the leading index.

Some approaches, i.e. the diffusion indexes by Torda (1985) and Chaffin and Talley (1989) are based on the percentage of individual indexes which decreases (or increases). However, here as in most papers, a leading composite index which is a weighted average of individual leading indicators will be used. Important

work on choice of individual indexes and their weights has been done.

3.2 Relations between earlier methods and surveillance

A turning point might be considered as a change from a phase of recession to a phase of expansion (or vice versa). Each phase might be modeled. Thus, general surveillance methods for detection of change from a model might be applied. Harrison and Veerapen (1994) discuss the use of CUSUM, sequences of SPRTs and a Bayesian decision rule to continually question model adequacy. Garbade (1977) describes the need (and previous ignorance) of methods to detect a lack of stability of the coefficients in econometric models. One of the methods discussed, namely that of Brown, Durbin and Evans (1975), is a CUSUM test of certain residuals. It is thus suited for running decisions. However, the simulation results given by Garbade (1977) concern a single decision for a fixed set of observations and not for repeated decisions for an ongoing process. There has later been further development of methods of this type, see, e.g. Hackl and Westlund (1991).

The CUSUM method as suggested by Brown, Durbin and Evans (1975) is not identical to the CUSUM method of quality control by Page (1954). As was mentioned above the latter one is advocated in a setting of repeated decisions while the former one is advocated in the setting of one test of hypothesis. Besides of that the limits in each sub-comparison are not the same. The former method has limits which aim to give the same significance level in the each subtest while in the latter the marginal false alarm rate conditioned that no earlier subtest was significant is an increasing function of time. The former approach is optimal with respect to minimal expected delay of alarms for a fixed false alarm rate only if the intensity of changes is very high in the beginning of the

series. Frisén and Wessman (1999) demonstrated that the predicted value of an alarm is very low in the beginning if the intensity is constant (which is usually assumed in the econometric applications). In that case the early alarms have little value as they cannot be trusted.

Fluctuation tests, which are CUSUM-type tests based on the changes in parameter estimates, are evaluated for some fixed lengths of the time series by Garbade (1977). Andrews (1993) was critical against them because the conditional significance levels are not constant. As was mentioned above, this is a drawback if each subtest is to be interpreted as a separate significance test - but not with a monitoring interpretation. However, another criticism was that the tests are not based on sufficient statistics.

Chu et.al.(1996) advocate monitoring methods which have a fixed (asymptotic) probability of any false alarm during an infinite long surveillance period. For some applications this might be important because a strict significance test is the goal. In that case, ordinary statements for hypotheses testing can be made. However, the price is high. For the case of no change, we have:

$$\lim_{t \to \infty} P(t_A \leq t) = \lim_{t \to \infty} \sum_{i=1}^{t} P(t_A = i| \; t_A \geq i) \, P(t_A \leq i) = \alpha < 1$$

$$\Rightarrow \lim_{i \to \infty} P(t_A = i| \; t_A \geq i) = 0$$

since

$$\alpha > \lim_{t \to \infty} P(t_A \leq t) \sum_{i=1}^{t} P(t_A = i| \; t_A \geq i)$$

This means that the probability to make an alarm a long time after the monitoring has started is very low. This in turn implies that the probability to make an alarm if a change occurs a long time after the monitoring has started also is very low. As was seen in Section 2.2 this monitoring approach is not an optimal solution

with the utility functions usually used at monitoring. It does not minimize the expected delay from the change to the alarm.

An important line of research is the nonlinear time series modeling approach with, e.g. STAR-models by Teräsvirta. A smooth transition between phases is assumed in contrast to earlier two-regime switching models. In the paper by Teräsvirta and Anderson (1992) you can see that even though the modeling fits well to a data set of quarterly logarithmic industrial production indexes for 13 OECD countries and "Europe" from 1960(i) to 1986(iv) the residuals are largest around the turning-points which confirms that the exact behavior around the turning-points is hard to model.

Many different tests of change at a pre-specified time are suggested in the econometric literature. They are usually in some way based on a likelihood-ratio or some asymptotically equivalent statistic. In the present kind of problems the time of change is unknown. A combination of tests by some technique of multiple statistical inference is sometimes used. Andrews (1993) suggests that the union intersection principle should be used. This implies that the maximal value of the test statistic (with respect to the value of τ) is used. The test statistic agrees with the general definition of the CUSUM statistic

$$\max_{\tau} LR(\tau)$$

used by, e.g. Siegmund (1985).

The two methods by Zarnowitz and Moore (1982) are explicitly stated as sequential signal systems. Their methods are multivariate (with Y(s) and a one-dimensional index L(s)) but does not utilize earlier information (before the decision time s) otherwise than by a rule of "natural ordering" of statements. Three different signals P1, P2 and P3 are possible for a peak and three signals T1, T2, and

T3 for a trough. The methods were defined by a set of verbal conditions. Here a description by graphs will be used. According to the first method, the level method, the signal is turned on if the observation (Y(s), L(s)) occurs in a new area of the Figure 1 and also the direction of the change into the area is the natural one according to Figure 2. The signal is turned off and a false alarm registered if the change is into an area which is not in the natural order. In the paper there is no interpretation of the quite possible change two or three steps forward (e.g. from the P1-area to the P3-area directly). The second method, the band-method, has different limits for the signal-areas as illustrated in Figure 3. Observe that there is an overlapping of P1- and T1-signals. Except for the signalling-areas there are also no-signal areas (unmarked in Figure 3). The rules for turning on or off signals are corresponding to those of the level method except for the different borders of the areas and that changes into or out of the no-signal areas are ignored. In Figure 4 and 5 their data (Table 4 columns 1 and 2 except for an obvious error in the sign of the Y-value of May 1980) are recorded in my kind of diagrams for the band and the level methods.
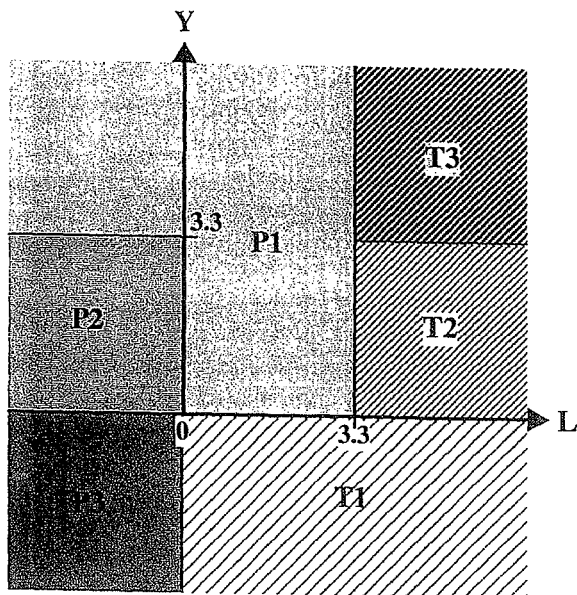
Figure 1. The level method by Zarnowitz and Moore (1982). Different signals are given according to in which area the index Y of industrial production and the leading index L are positioned at the time of decision.
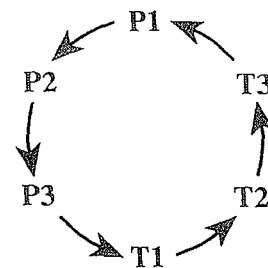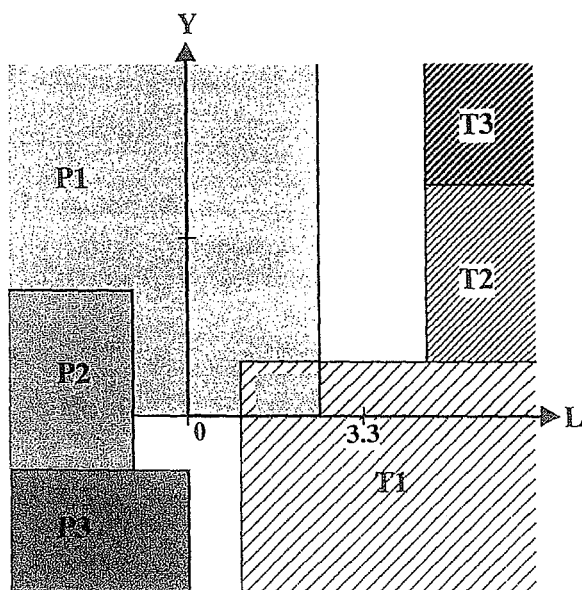

Figure 2. The "natural" order of signals.

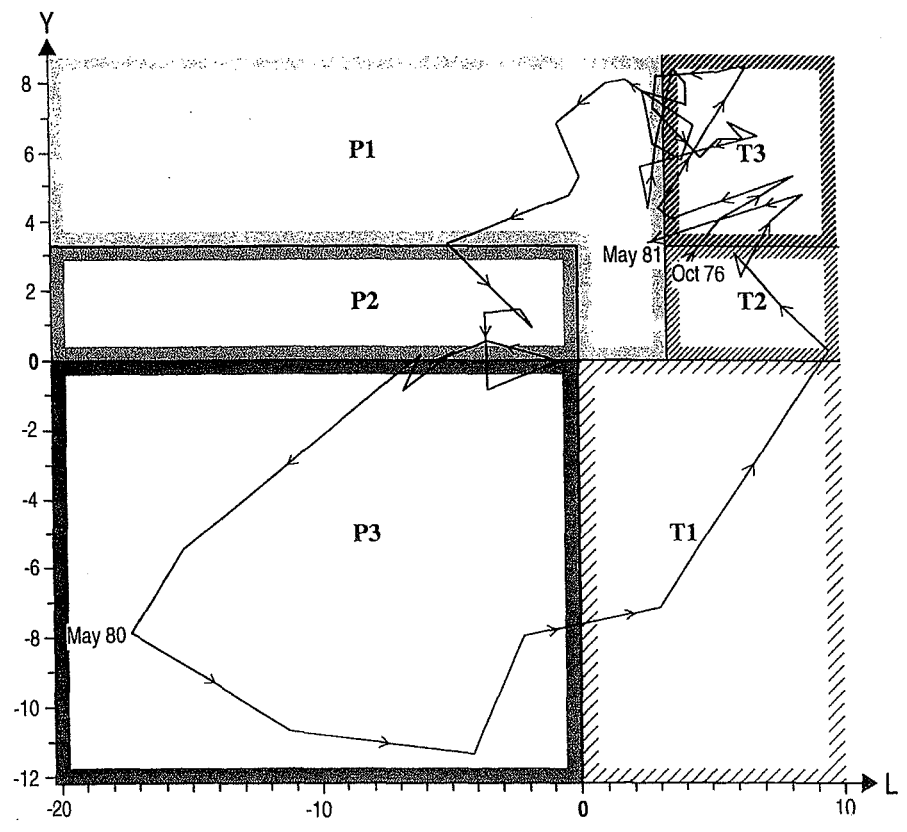

Figure 3. The band method by Zarnowitz and Moore.

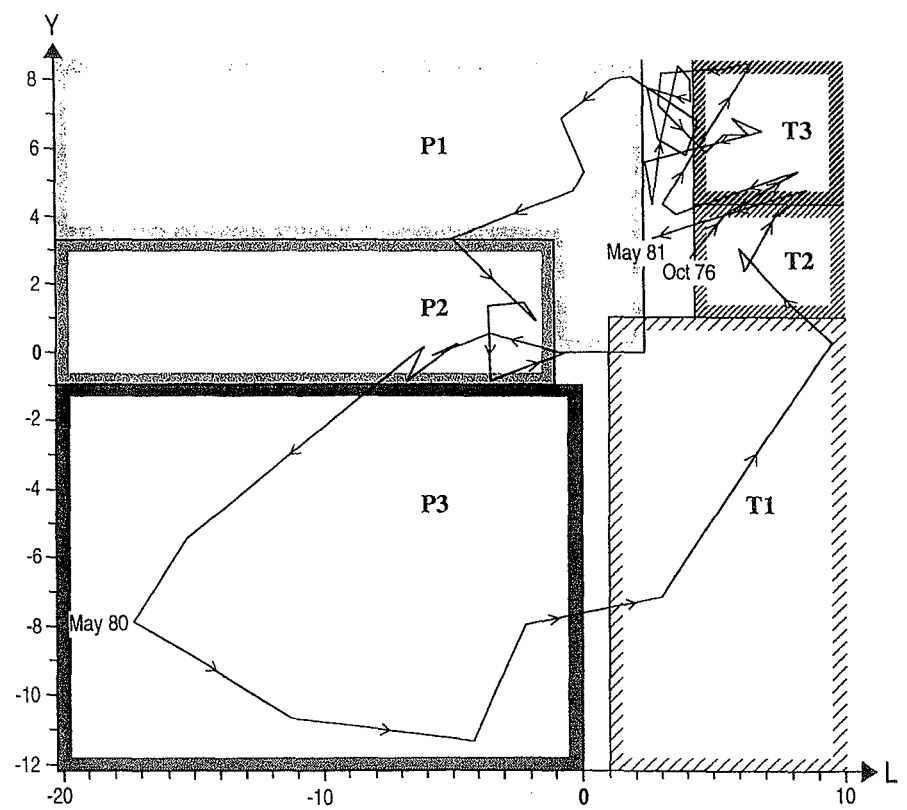**Figure 4.** The level method and data from U.S.A.

Figure 5. The band method and data from U.S.A.

A modification of the methods by Zarnowitz and Moore was suggested by Niemira (1983). However this method uses information at time s which is not available at time s but later. This is a problem also encountered when final revisions of an index are used. As an example, the Swedish final index of industrial production might be as old as 32 months. Diebold and Rudebusch (1991) analysed this problem with the help of a new data set which contains every preliminary, provisionally revised, and final estimate of the Department of Commerce composite index of leading indicators.

Nazmi (1993) based a signal system on a probit-function of a linear expression in leading indicators and a cyclical variable. This statistic is then "filtered" to avoid too many false alarms. Three filters were examined. They are of the type: signal if three values in a row of the statistic exceed a certain limit. The use of these filters is a way to take into account values of the statistic not only for the last time point but also for earlier ones. The way this is done is equivalent to what in quality control is named "Shewhart with warning lines". The filter is also equivalent to that of the "sets" method in the literature on medical surveillance.

Different ways of taking care of seasonal effects will not be analysed here. However it is clear that a transformation

$$L'(t)=L(t)- \sum_{i=t-12}^{t-1} L(i)/12$$

which is often used (e.g. Zarnowitz and Moore 1982), will destroy some of the power of detection of a change. The technique by Diebold and Rudebusch (1991) not to use the data from the last year when estimating the seasonal effects will give a better chance to detect recent changes. Sometimes transformations which utilize

observations after the decision time are used. Such methods cannot be used in a sequential signaling system.

The Bayesian modeling approach to econometrics and structural change is described, e.g. in the book by Broemeling and Tsurumi (1986). A good review of this line of research is given by Tsurumi (1988). Posterior distributions are used for turning point detection by, i.e. Neftci (1982), Hamilton (1989), Diebold and Rudebusch (1991), Webb (1991), Jun and Joo (1993) and Birchenhall et al. (1999). Jun and Joo use a random shock model, Birchenhall et al. use a logistic model and the others use different models for different phases. The models are estimated by the data sets. Stable models and large samples are required to get good estimates of the posterior distributions. The new method suggested in Section 3.4 is also based on the posterior distribution but does not require estimates of parameters.

3.3 Illustration by Swedish data

The lags between monthly data of an index of industrial production and an index of leading indicators (during the period January 1960 to April 1987) in Sweden were examined. The best linear relation as measured with the coefficient of determination R^2 (=0.50) was obtained by a lag of 10. The relation is illustrated in Figure 6.
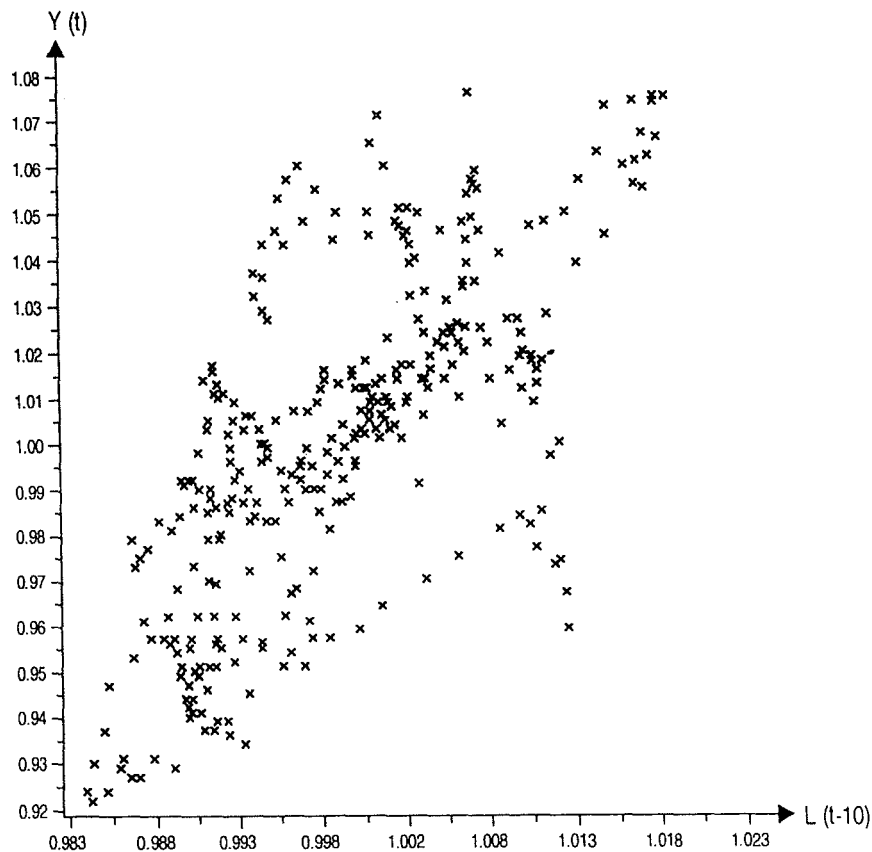

Figure 6. The relation between the index of industrial production, Y, in Sweden and the leading index, L, with a lag of 10 months.

In Figure 7 a, b and c Swedish data for different periods are recorded in the same way as the data in Figure 5 to suit the same methods. As can be seen there is no way of positioning alarm-lines according to the techniques of Zarnowitz and Moore (1982) that would give reasonable results. One explanation to this might be that it is not the values but the very turning of the leading indicator which gives indication of a turn in the industrial production.
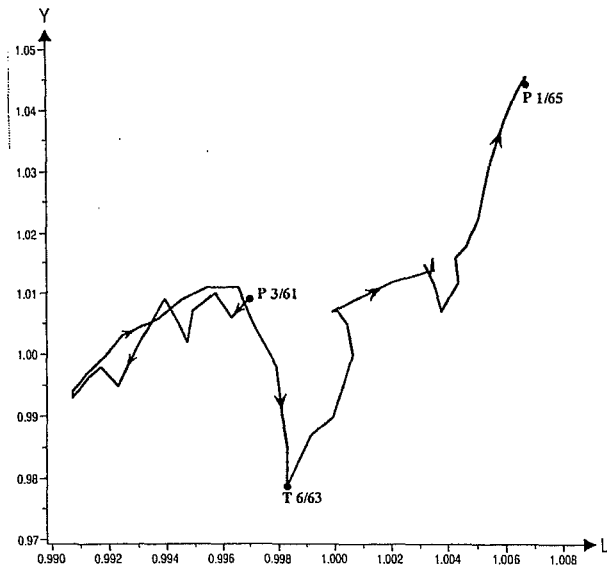
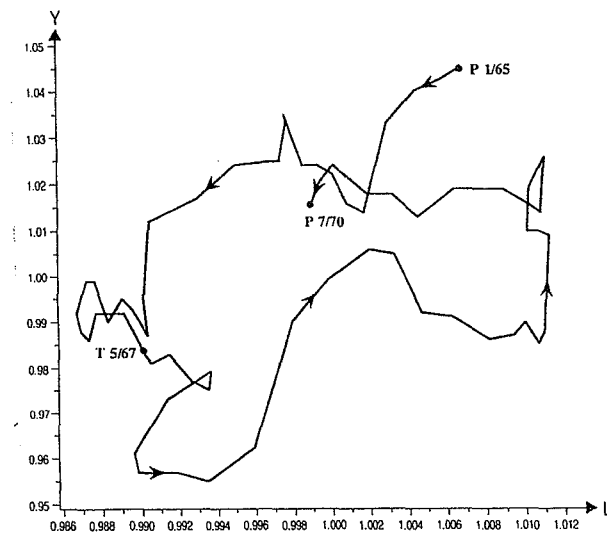**Figure 7a.** The period between the peak in Mars 1961 to the peak in January 1965 with a trough in June 1963.

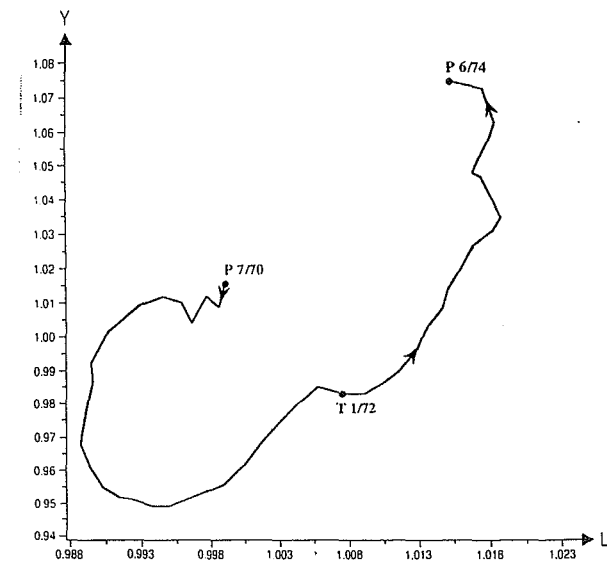**Figure 7b.** The period January 1965 to July 1970.

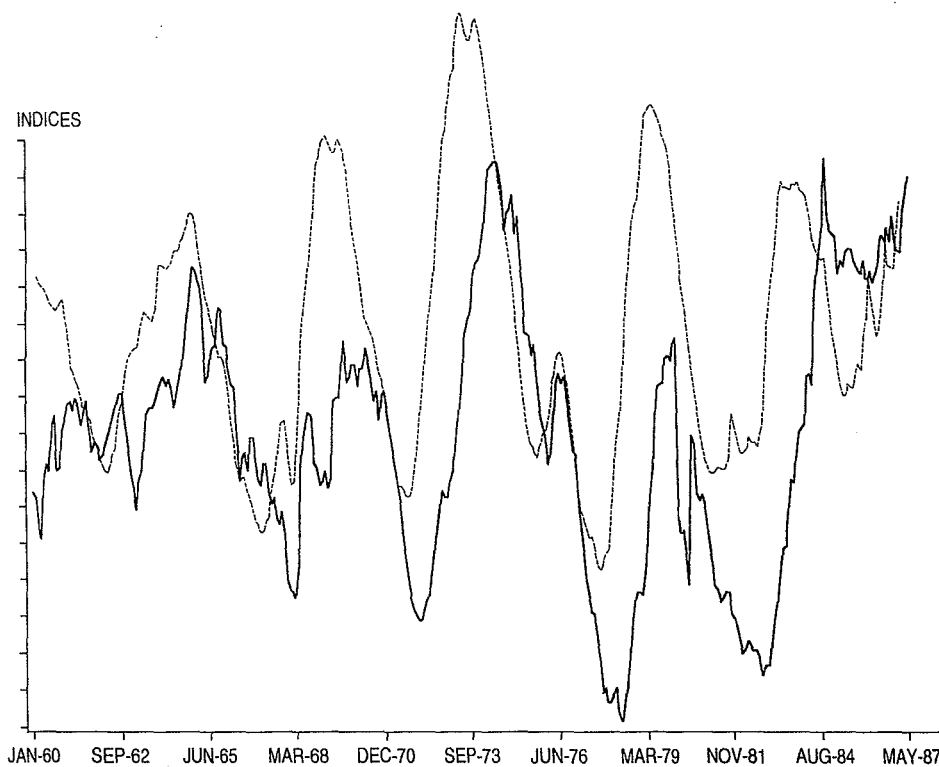**Figure 7c.** The period July 1970 to June 1974.

Figure 8a. The index of industrial production (solid curve) and the leading index (dotted curve) in Sweden January 1960 to April 1987.
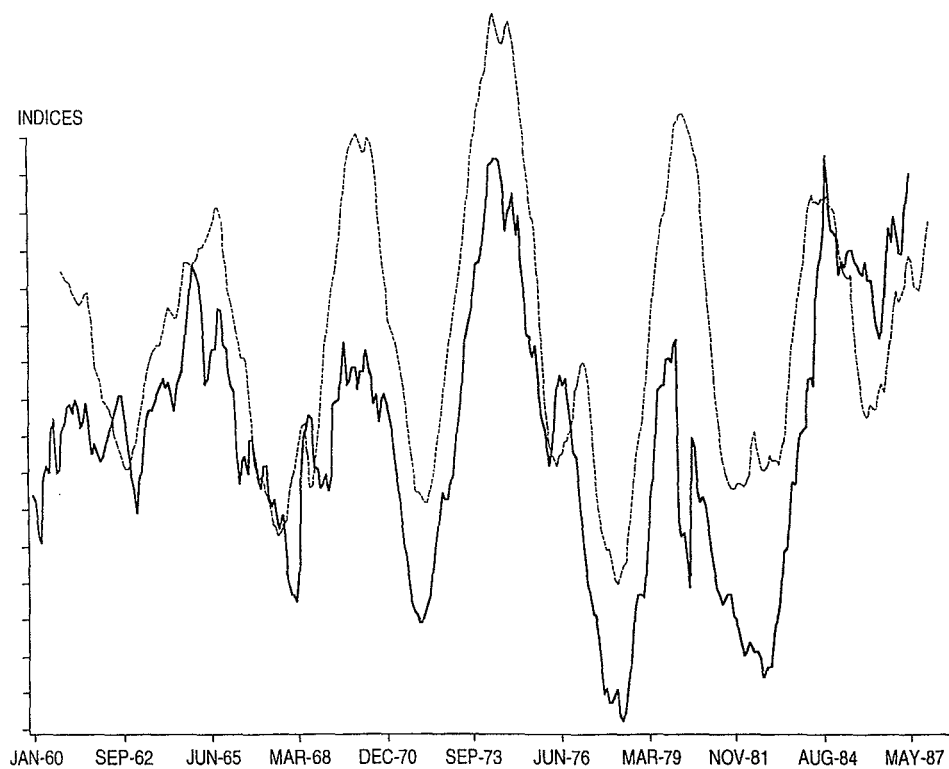

Figure 8b. The same data as in Figure 8a but the leading index is lagged 10 months.

3.4 A new nonparametric method based on change in monotonicity

As the term "turning-point" indicates, and is explicitly claimed by, i.e. Nazmi (1993), it is the monotonicity and not the absolute values which characterize the changes to be detected. This is confirmed by Swedish data in Figures 8a and 8b. In Figure 8a the indexes of industrial production and a leading index from January 1960 to April 1987 are given with a common time axis (and arbitrary axis for the indexes). In Figure 8b the same data are reproduced with the leading index lagged 10 months. As can be seen, the turning of the leading index is a good indicator that the industrial production will turn about 10 months later. The levels of the indexes seem however to have little impact. This might be the reason why the Zarnowitz Moore technique did not work well when used in Section 3.3. Contrary to most earlier methods, the new method suggested here is not based on the level of the indexes but on the monotonicity.

There is much debate about the proper way of making models for business cycles and leading indexes. For example, there are different opinions on the symmetry around the turning-points (see e.g. Westlund and Öhlén 1991). Westlund (1989) discussed the robustness of methods. Here only the change in monotonicity from increasing values to decreasing (or vice versa) at a turning point will be used. By concentration on the known qualitative properties, robustness is achieved.

According to Frisén and de Maré (1991) optimal methods to discriminate between the states

$$C(s) = \{s\text{-}d < \tau \leq s\} \quad \text{and} \quad D(s) = \{\tau > s\},$$

where s is the time of decision, τ is the change point and d is a constant, will be based on the likelihood ratio. As was seen in Section 2.2 this optimal LR method has an alarm set consisting of those outcomes X for which the likelihood ratio exceeds a limit

$$f_X(x_s \mid C) \,/f_X(x_s \mid D) = p(x_s) > k_s$$

where k_s does not depend on the data but on s. This is identical to the requirement that the posterior distribution exceeds a constant K. We have

$$p(X_s) > \frac{P(\tau > s)}{P(s - d < \tau \le s)} \; \frac{K}{1-K}$$

The time period (s-d, s) is the period specified by C in which it is important to detect a turn. If d=s+1 the method will be optimal to detect all turns before s.

As pointed out by Ghysels et.al. (1998), most studies in the vast literature on testing for structural change have paid most attention to the linear regression model. They studied a nonlinear dynamic model for a fixed period. For the case of successive decisions with increasing number of observations the most studied case of monitoring is the case of independently normally distributed variables and with a shift in mean from one value, say zero to another value, say M. If the variance is constant over time, the scale can be chosen to make the variance equal to one. In this case we have

$$p(x_s) = \sum_{k=s-d}^{s} \frac{\pi_k}{P(s - d < \tau \le s)} \exp\left\{ \frac{1}{2} \sum_{u=k}^{s} [(X(u)-0)^2 - (X(u)-M)^2] \right\}$$

This case, with d=s, was studied in detail by Frisén and Wessman (1999). Now we have another situation since the mean is not changing from one constant value to another but the monotonicity is changing. In this case, we do don't have two completely specified distributions to compare. We can thus not construct a uniformly optimal likelihood ratio method. Instead, a maximum likelihood ratio method under the different conditions of monotonicity is suggested. Now we have

$$p(x_s) = \sum_{k=s-d}^{s} \frac{\pi_k}{P(s-d<\tau \le s)} \exp \frac{1}{2}\{Q(0) - Q(k)\}$$

where Q(k) is the quadratic deviation from the best model with a turn at time k and Q(0) is the deviation from the best model without a turn in the specified time period. The deviations are based on the first s observations. These deviations can be calculated by the algorithms and computer programs given in Frisén (1980) and (1986).

As was demonstrated by Frisén and Wessman (1999) the method which is optimal for a specified distribution can often be well approximated by a method without these weights. When we have

$$p(x_s) = \sum_{k=s-d}^{s} \exp \frac{1}{2}\{Q(0) - Q(k)\}$$

Even though the conditions of independent normally distributed observations might not be exactly fulfilled for the economic time series at hand the alarm-function can be expected to be powerful if not optimal under the true conditions. Often (e.g. Diebold and Rudebusch 1991) independently normally distributed variables are assumed when likelihoods are computed. By transformation of the original data or by introducing the dependency it will be possible to improve the efficiency of the method. General methods for surveillance of dependent data have been suggested by e.g. Liu and Tang (1996) and the techniques might be used also for this case. The method suggested in this section is nonparametric with respect to regression functions since only monotonicity properties are used. However, it is not nonparametric with respect to distribution. The estimates used to compute the deviations, are ML estimators only at normal distribution. Otherwise, they are LS estimators.

4. EVALUATIONS

4.1 General evaluations of methods of surveillance

Most evaluations of methods for forecasting of turning-points in business cycles are made by one suggestive example. However, recently large scale comparisons by application to several data sets are made by, e.g. Webb (1991) and Diebold and Rudebusch (1991). In this section theoretical measures of the goodness of methods for continual surveillance will be reviewed.

The usual measures of a test's performance, namely the significance level and the power, would have to be generalized in any of many possible ways to take into account the dependence on the length of the period of surveillance and the time point τ where the change occurs. Frisén (1992) gives suggestions of such measures and demonstrates that important features of methods will be revealed.

In quality control the average run length (ARL) until an alarm occurs is often used. See e.g. Wetherill and Brown (1990). The average run length, ARL^0, is the average number of runs until an alarm occurs when there is no change in the system under surveillance. The average run length under the alternative hypothesis, ARL^1, is the mean number of decisions that must be taken to detect a true level change (that occurred at the same time as the inspection started).

The distributions of the run length (RL) contain the information necessary for an evaluation of a method or a comparison between methods. The actual comparison is usually based on the average run length, but also the median or some other percentile could be considered. The run-length distributions, especially those

connected with the alternative hypotheses, are usually skew. This was illustrated by, e.g. Frisén and Wessman (1999). Only one summarizing measure of the distribution is thus not enough.

Unfortunately, the forecasting of business cycle turning points sometimes produces false signals. The distribution when the process has not changed the regime can be described by a measure α_t which corresponds to the probability of erroneous rejection of the null hypothesis, the level of significance, but is a function of the time t. α_t is the probability of an alarm no later than at t given that no change has occurred.

$$\alpha_t = P(RL \leq t \mid D(t)).$$

The distance between the change and the alarm, sometimes called "residual RL" (RRL) is of interest in many cases. The optimality conditions by Girshick and Rubin (1952) and Shiryaev (1963) are based on this distance. One characterization of the distribution of the RRL is the probability that the RRL is less than a certain constant d (the time limit for successful rescuing action). This measure, PSD(d), the probability of successful detection, is the probability to get an alarm within d time units after the change has occurred, conditioned that there was no alarm before the change. The PSD is a function of the time distance d, the time of the change t' and μ^1.

$$PSD(d, t', \mu^1) = P(RL < t' + d \mid RL \geq t')$$

The predictive value $PV(s) = P(C(s) \mid A(s))$ is the relative frequency of motivated alarms among all alarms at a certain point of time. This measure is a function of the incidence q, and the time t_A of the alarm. It gives information on whether an alarm is a strong

indication of a change or not. The trust you should have in an alarm is measured by the predictive value. It gives the balance between the false and the motivated alarms. If the predictive value is constant over time, you can interpret an alarm in the same way whenever it happens.

The difference between passive and active surveillance (as discussed in Section 3.2) affects the different possible error rates as can be seen in Frisén and de Maré (1991). It will also have consequences for the predicted value and the posterior distribution.

A method based on the posterior distribution, PD, has the alarm set

$$A(s) = [X_s; PD(s) > c].$$

At passive surveillance, that is when our actions at an earlier time point do not affect the distributions this implies $PV(s) > c$, that is a method based on the predictive value. Typically PV increases to one when s increases. At active surveillance, when the surveillance will be stopped as soon as an alarm occurs, none of this holds. Instead, typically PV has an asymptote below one and PV is not monotonically increasing for all methods.

Simultaneous visual illustration, of the measures of goodness described above for different situations and methods, is described by Frisén and Cassel (1994). The self-instructing computer program is available free of charge from the author.

4.2 An example of evaluation by a data set on business cycles

Often the performance of a new method is illustrated by one successful application. This does not make the evaluations suggested above superfluous. A thorough evaluation of a statistical method can never be based on just one realization of the stochastic

variables, particularly if the outcome is studied before the design of the method. The two ways of evaluation, by the stochastic properties of the method and by application to real data, complement each other.

In this section an example of a detailed evaluation will be given. The method by Zarnowitz and Moore (1982) and their data will be studied since this paper is one where the sequential nature of the problem is recognized and the paper is often cited. It has also later inspired several similar methods and evaluations. Even though the paper is very well written, with a detailed concern about many problems in the area, there are of course some problems which could have been treated more in detail. Here some detailed evaluations of their level method are made in order to illustrate the need of very precise statements on the interpretations and properties of signalling systems of this kind.

In the paper there is a statement on page 57 that all peaks and troughs are identified with a minimum of delays and false alarms. The statement is supported by a table where the signals near each peak or trough are given and by a discussion. Here some details concerning the meaning of "false alarms", "minimum delay" and "identification" of peaks are given. To illustrate the discussion, all peaks and troughs and all signals during the time period October 1976 to May 1981 are marked on a time axis in Figure 9.
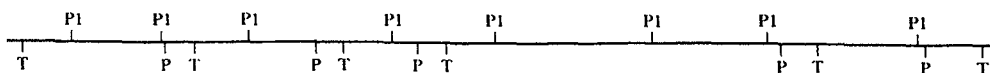
Figure 9. All peaks (P), troughs (T) and P1-signals with the level method.

Possible interpretations of "identified" could be:

1. The absolute (positive or negative) distance from each true peak (P) to the nearest P1-signal is less than a distance d. In that case d has to be 2 years for the claim that all peaks were identified to be true.

2. The distance after each peak until the next P3-signal comes should be less than 4 months. All but the last peak was identified in this respect.

Clearly these two kinds of "identification" should be interpreted quite differently and clarification is necessary.

Possible interpretations of "minimum delay" could be:

1. The P-signal comes after the peak but not more than 4 months.

2. The P-signal comes well in time before the peak. The distance between the P1-signal and the next peak varies between 1 and 91 months (median 18 months). For the P3-signal the time until the cycle turns up again is between 6 and 10 months.

The different measures of delay demonstrate the difficulties to know which action is appropriate at a signal.

Possible interpretations of "false alarm" could be:

1. The distance between a P1-signal and the next peak is more than, say, 19 months. In that case 4 out of the 7 P1-alarms are false.

2. There has not been a peak within 0 - 4 months before a P3-signal.

The difference in interpretation of the different signals is demonstrated by their different timing with respect to the peak they are supposed to predict

or confirm. In Figure 10 the distance between the time t(P1) of a P1 signal and the time t(Pb) of the peak before the signal is given. It is related to the distance between the time t(Pa) of the peak after and t(P1). The figure illustrates the difficulties of interpretation of the P1 signal. The conclusion by Webb (1991) that it is hard to forecast more accurately than by using an uninformative naive indicator seems relevant also here.
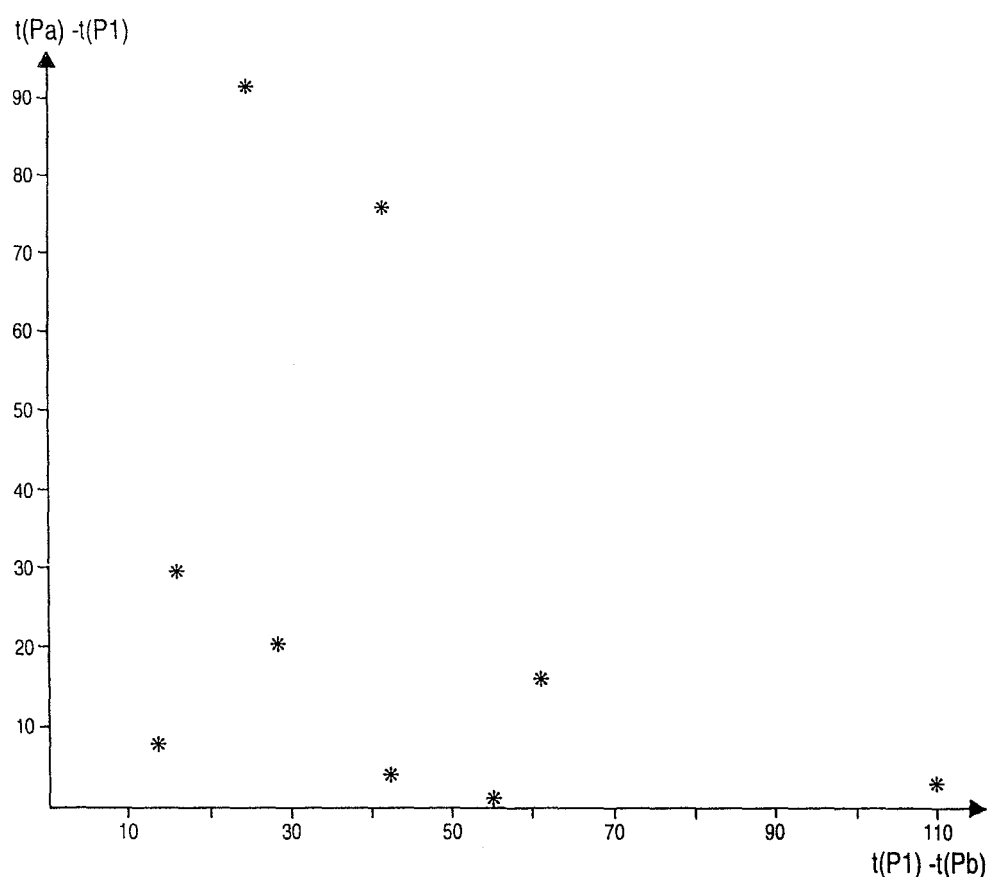

Figure 10. Illustration of the relation between the distances from the P1 signal to the peak after and the peak before the signal.

The P3-signal tends to come after the peak and near the phase of expansion. The action that should be triggered by this or the P2- signal seems unclear. When the P2-signal comes in an excessive phase in their data set, the phase continues for 0 - 34 months. When the P2-signal comes in a recession phase, it ends in less than 14 months.

All limits above are determined to be as favourable as possible with respect to the observed data. Also other methods where limits are set to suit a certain data set will tend to appear too favourable. The difference between apparent and actual error rate (Efron 86) used in the general theory on predictions is relevant also here. Nazmi (1993) used one period for the estimation of parameters and another one for the evaluation to avoid this problem.

5. CONCLUDING REMARKS

An important distinction is that between a single decision and a sequence of decisions. At a single decision alarm for a great value of the posterior probability or (when there is no prior) significance at an ordinary test is natural. For a sequence of decisions the characteristics of the sequence (such as constant predictive value or expected waiting time between change and alarm) become interesting. Most papers on detection of turning points in business cycles are concerned with the testing of hypotheses. However, it might also be of value to focus on the sequential aspects of economic policy decision making and the need for optimal timing of economic policy programs as has also been advocated by Niemira (1991).

As was seen in Section 2.2 optimal surveillance is based on likelihood ratio statistics. Many of the methods suggested in the literature on business cycles forecasting are based on statistics which are (at least asymptotically) equivalent to likelihood ratios. Important differences are the specifications of the different states of the process used in the ratio. Another important difference is how the ratios are combined to form a method of testing a hypothesis or a method of sequential signals.

The statistics for different times are combined to form a signaling system. This is sometimes named filtering. The weight you give the last value of the statistic compared to earlier ones is crucial for the properties such as timeliness of the signalling system. Most of the signalling systems only consider the last value (as in the Shewhart method of quality control). This is optimal, if only change at the last time point is of interest, but might have reasonable properties also for other cases. Others consider also earlier values with equal or unequal weights. By comparing the weighting system with the optimal ones it is possible to tell for which cases the methods will work well.

Neftci (1982) and Jun and Joo (1993) base their methods on the posterior distribution for a turning-point and their methods are thus related to the optimal nonparametric method suggested above. Also their methods satisfy the optimality condition by Shiryaev (1963) if their assumptions on models are true and if the parameters are known. Long series of data and stable models are important for the estimation in their methods in contrast to the non-parametric approach suggested here.

The states of the systems in the likelihood ratio are those between which discrimination is needed. It is important to decide on which the main characteristics of the change are and how strong assumptions of the models you are willing to make. If the only safe characteristic of the change is the change in monotonicity then the alarm-function should be based on this if you are concerned about robustness. The non-parametric approach described here avoids the need of long data sets and stable models necessary for the estimation in parametric methods.

The evaluation should correspond to the aims. The detailed evaluations above demonstrate the need for extremely precise statements. It is necessary to give a clear interpretation of an alarm

signal in order to be able to evaluate whether the signal is useful for the specified interpretation. There is a great difference between a signal (P1) warning of a possible change to recession after a long time and a signal (P3) indicating that the recession is already in effect.

The problem of detection of change of phases of business cycles is important, and it is necessary to specify the aims and then to design a method which meets the aims. You have to know how much you can rely on it and how to interpret an alarm. Several ways of evaluations, as described above, have to be used.

School of Economics and Commercial Law, University of Göteborg

REFERENCES

Andrews, D.W. , 1993. Tests for parameter instability and structural change with unknown change point. Econometrica, 61, 821-856.

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

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

Broemeling, L.D. & Tsurumi, H., 1986. Econometrics and structural change. New York: Marcel Dekker.

Brown, R.L., Durbin, J. and Evans, J.M., 1975. Techniques for testing the constancy of regression relationships over time, Journal Roy. Stat. Soc. Ser. B, 37, 149-192.

Chaffin, W.W. & Talley, W.K., 1989. Diffusion indexes and a statistical test for predicting turning points in business cycles. International Journal of Forecasting, 5, 29-36.

Chu C., Stinchcombe M. & White H., 1996. Monitoring structural change. Econometrica, 64, 1045-1065.

Diebold, F.X. & Rudebusch, G.D., 1989. Scoring the leading indicators. Journal of Business, 62, 369-91.

_____ 1990. A nonparametric investigation of duration dependence in the American business cycle. Journal of Political Economy, 98, 987-1107.

_____ 1991. Turning point prediction with the composite leading index: An ex ante analysis. in Lahiri, K. & Moore, G.H. (eds), Leading economic indicators. New approaches and forecasting records. Cambridge: Cambridge University Press.

Edlund, P. & Sogaard, H.T.,   1993. Fixed versus time-varying transfer functions for modelling business cycles. Journal of Forecasting, 12, 345-364.

Efron, B.,   1986. How biased is the apparent error rate of a prediction rule? Journal of the American Statistical Association, 81, 461-470.

Frisén, M.,   1980. U-shaped regression. in   Compstat. Proceedings of computational statistics.

_____ 1986. Unimodal regression. The Statistician, 35, 479-485.

_____ 1992. Evaluations of methods for statistical surveillance. Statistics in Medicine, 11, 1489-1502.

Frisén, M. & Cassel, C.,   1994. Visual evaluations of statistical surveillance. Research report, Department of Statistics, University of Göteborg.

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

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

Garbade, K.,   1977. Two Methods for Examining the Stability of Regression Coefficients. Journal of the American Statistical Association, 72, 54-63.

Ghysels E., Guay A. & Hall A.,   1997. Predictive tests for structural change with unknown breakpoint. Journal of Econometrics, 82, 209-233.

Girshick, M.A. and Rubin, H.,   1952. A Bayes approach to a quality control model. The Annals of Mathematical Statistics, 23, 114-125.

Hackl, P. & Westlund, A.H.,   1989. Statistical analysis of structural change. An annotated bibliography. Empirical Economics, 14, 167-192.

_____ 1991. Economic Structural Change: Analysis and forecasting. Springer.

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

_____ 1993. Estimation, inference and forecasting of time series subject to changes in regime. in Maddala, G.S. & Rao, C.R. & Vinod, H.D. (eds), Handbook of statistics: Econometrics. Amsterdam: North-Holland.

Harrison, P.V. & Veerapen, P.P., 1994. A Bayesian decision approach to model monitoring and cusums. Journal of Forecasting, 13, 29-36.

Hunter, J., 1986. The Exponentially Weighted Moving Average. Journal of Quality Technology, 18, 203-210.

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

Krämer, W., Ploberger, W. & Alt, R., 1988. Testing for structural change in dynamic models. Econometrica, 56, 1355-1369.

Lahiri, K. & Moore, G., 1991. Leading economic indicators: New approaches and forecasting record. Cambridge: Cambridge University Press.

Lee, K. and Shields, K. (2000) Expectations Formation and Business Cycle Fluctuations: An Empirical Analysis of Actual and Expected Output in UK Manufacturing, 1975-1996. Oxford Bulletin of Economics and Statistics, 62, 463 - 490.

Liu, R. Y. and Tang, J. (1996) Control charts for dependent and independent measurements based on bootstrap methods. Journal of the American Statistical Association, 91, 1694-1707.

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

McCulloch, J.H., 1975. The Monte-Carlo Cycle in Business Activity. Economic Inquiry, 13, 303-321.

Moustakides, G.V., 1986. Optimal stopping times for detecting changes in distributions. Annals of Statistics 14, 1379-87.

Nazmi, N., 1993. Forecasting Cyclical Turning Points with an Index of Leading Indicators: A Probabilistic approach. Journal of Forecasting, 12, 215-225.

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

Niemira, M.P., 1983. Sequential signals of recession and recovery: revisited. Business Economics, 18, 51-53.

_____ 1991. An international application of Neftci's probability approach for signalling growth recessions and recoveries using turning point indicators. in Lahiri, K. & Moore, G.H. (eds), Leading economic indicators. New approaches and forecasting records. Cambridge: Cambridge university press.

Nyblom, J., 1989. Testing for the constancy of parameters over time. Journal of the American Statistical Association, 84, 223-230.

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

Ploberger, W. & Krämer, W., 1990. The local power of the cusum and cusum of squares tests. Econometric Theory, 6, 335-347.

Pollak, M., 1985. Optimal stopping times for detecting changes in distributions. Annals of Statistics 13, 206-227.

_____ 1987. Average run length of an optimal method of detecting a change in distribution. Annals of Statistics, 15, 749-779.

Ritov, Y., 1990. Decision theoretical optimality of the CUSUM procedure. Annals of Statistics 18, 1464-1469.

Roberts, S.W., 1959. Control Chart Tests Based on Geometric Moving Averages. Technometrics, 1, 239-250.

_____ 1966. A comparison of some control chart procedures. Technometrics, 8, 411-430.

Robinson, P.B. and Ho, T.Y., 1978. Average Run Lengths of Geometric Moving Average Charts by Numerical Methods. Technometrics, 20, 85-93.

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

Shukur, G., 1993. Demand for Meat and Fluid Milk in Sweden 1960-1989. Test for Structural Change.. Diss., Department of Statistics, Lund.

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

Smith, A.F.M., West, M., Gordon, K., Knapp, M.S. and Trimble, M.G., 1983. Monitoring kidney transplant patients. The Statistician, 32, 46-54.

Stekler, H.O., 1991. Turning point predictions, errors, and forecasting procedures. in Lahiri, K. & Moore, G.H. (eds), Leading economic indicators. New approaches and forecasting records. Cambridge: Cambridge University Press.

Stock, J. H. & Watson M. W., 1994. Business cycles, indicators and forecasting. Chicago, University of Chicago Press.

Teräsvirta, T. & Anderson, H.A., 1992. Modelling nonlinearities in business cycles using smooth transition autoregressive models. Journal of Applied Econometrics, 7, 119-.

Torda, T., 1985. Purchasing management index provides early clue on turning points. Business America, 8, 11-13.

Tsurumi, H., 1988. Survey of Bayesian and nonbayesian testing of model stability in econometrics. in Spall, J. (eds), Bayesian analysis of time series and dynamic models.

Webb, R.H., 1991. On predicting the stage of the business cycle. in Lahiri, K. & Moore, G.H. (eds), Leading economic indicators. New approaches and forecasting records. Cambridge: Cambridge university press.

Westlund, A.H., 1984. Sequential moving sums of squares of OLS residuals in parameter stability testing. Quality and Quantity, 18, 261-273.

_____ 1989. Robust estimation and prediction of economic systems: the case of partial structural variability. Quality and Quantity, 23, 61-.

_____ 1993. Business Cycle Forecasting. Journal of Forecasting, 12, 187-196.

Westlund, A.H. & Zackrisson, U., 1986. On the prediction of structurally varying systems. Technological Forecasting and Social Change, 30, 63-72.

Westlund, A.H. & Öhlén, S., 1991. On testing for symmetry in business cycles. Empirical Economics, 16, 479-.

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

Zacks, S., 1980. Numerical determination of the distributions of stopping variables associated with sequential procedures for detecting epochs of shift in distributions of discrete random variables. Communications in Statistics. Simulations & Computations, 9, 1-18.

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

Zarnowitz, V., 1985. Recent work on business cycles in historical perspective: a review of theories and evidence. Journal of Economic Literature, 102, 523-80.

Zarnowitz, V. & Moore, G.H., 1982. Sequential signals of recessions and recovery. Journal of Business, 55, 57-85.

Öller, L., 1987. Stabiliseringspolitik och prognoser. Ekonomiska samfundets tidskrift, 40, 23-27.

_____ 1990. Forecasting the business cycle using survey data. International Journal of Forecasting, 6, 453-461.