



Research Report
Department of Statistics
Göteborg University
Sweden

**Statistical surveillance of cyclical processes
with application to turns in business cycles**

**Eva Andersson
David Bock
Marianne Frisé**

**Research Report 2002:8
ISSN 0349-8034**

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

Statistical Surveillance of Cyclical Processes with Application to Turns in Business Cycles

EVA ANDERSSON, DAVID BOCK AND MARIANNE FRISÉN*
Göteborg University, Sweden

ABSTRACT

On-line monitoring of cyclical processes is studied. An important application is early prediction of the next turn in business cycles by an alarm for a turn in a leading index. Three likelihood based methods for detection of a turn are compared in detail. One of the methods is based on a Hidden Markov Model. The two others are based on the theory of statistical surveillance. One of these is free from parametric assumptions of the curve. Evaluations are made of the effect of different specifications of the curve and the transitions. The methods are made comparable by alarm limits, which give the same median time to the first false alarm, but also other approaches for comparability are discussed. Results are given on the expected delay time to a correct alarm, the probability of detection of a turning point within a specified time and the predictive value of an alarm.

KEY WORDS monitoring; optimal; likelihood ratio; Hidden Markov Model; non-parametric

Short title: Statistical Surveillance of Cyclical Processes

*Correspondence to: Marianne Frisén, Department of Statistics, P.O.Box 660, SE-40530, Sweden, phone +46 317731255, fax +46 317731274, Marianne.Frisen@Statistics.GU.se

1. INTRODUCTION

On-line monitoring of cyclical processes is important in different areas. An important application is early prediction of the next turn in business cycles by an alarm for a turn in a leading index and this case will be the starting point for this paper. Another application in a quite different area is natural family planning. Temperature and other cyclical leading indicators are used to predict the ovulation.

The turn in a business cycle is a change from a phase of recession to one of expansion (or vice versa). Warning for a turn can be made by using information from one or several time series, which are leading in relation to the actual business cycle. By applying a system for detection of the turning points of a leading indicator we can receive early indication about the future behavior of the business cycle. For reviews and general discussions see e.g. Neftci (1982), Zarnowitz and Moore (1982), Westlund and Zackrisson (1986), Hackl and Westlund (1989), Diebold and Rudebusch (1989), Hamilton (1989), Zellner et al. (1991), Jun and Joo (1993), Lahiri and Wang (1994), Li and Dorfman (1996), Koskinen and Öller (1998) and Birchenhall et al. (1999).

In recent years methods based on posterior probability or likelihood have been in focus. There are proofs for their optimality properties in the general theory on statistical surveillance (see e.g. Shiryaev (1963) and Frisé and de Maré (1991)). In this report, the effects of different specifications of likelihood-based systems for detection of turning points are examined.

The performances of three methods for turning point detection in leading indicators are compared in detail. All three methods are based on the likelihood, but there are differences in model specifications, how much information that is used and parameter estimation. The HMLin method is based on a regime switching Hidden Markov Model (HMM) and agrees with a model which is piecewise linear as will be discussed in Section 2.1. It is similar in several aspects to e.g. the method presented by Koskinen and Öller (1998). HMM is suggested for business cycle modeling and prediction by Hamilton (1989) and is used by e.g. Lahiri and Moore (1991), Lahiri and Wang (1994), Layton (1996), Koskinen and Öller (1998) and Gregoir and Lengart (2000). The SRLin method is derived here by the Shiryaev-Roberts (SR) technique under the assumption of a piecewise linear model. The SRnp method was suggested by Frisé (1994) and evaluated by Andersson (2001) and Andersson (2002) under the name of MSR since it is a Maximum likelihood version of the SR method. The SRnp method is a non-parametric version of the SRLin method with no parametric assumption on the shape of the curve.

Here, simulation studies are made to evaluate and compare the three methods. Special concern is given to the different ways to avoid false signals, to utilize prior information and the effect of assumptions regarding the shape of a turning point and the distribution of the time of transitions. Comparison of the three methods by application to Swedish data is made by Andersson et al. (2002).

The inference situation can be described as one of surveillance, since we have continual observation of a time series with the goal of detecting the turning point in the underlying process as soon as possible. Repeated decisions are made, the sample size is increasing and no null hypothesis is ever accepted. Thus, the inference situation is different from that with a fixed number of observations. For general reviews on statistical surveillance, see Shiryaev (1963), Frisé and de Maré (1991), Wetherhill and Brown (1991), Srivastava and Wu (1993), Lai (1995), Frisé and Wessman (1999) and Frisé (2002).

The performance is evaluated using measures such as expected delay of a signal, probability of successful detection and predictive value of an alarm, as suggested by Frisé (1992). This is related to the evaluation of the chronology used by e.g. Kontolemis (2001) but quite different to the MSE or AIC of errors in forecast values used by e.g. Bidarkota (2001). Some approaches, like that of Birchenhall et al. (1999), discuss both forecasting of values and detection of a regime change in a leading index. Here we deal solely with the detection of a change.

Section 2 contains a description of different likelihood based approaches, specifically of the HMLin, the SRLin and the SRnp methods. It also contains theoretical analyses of the effect of some assumptions. In Section 3 results from a simulation study on the effects of different methods and different assumptions are presented. The choice of models for the simulation study is motivated by similarity to Swedish data. Section 4 contains a summarizing discussion.

2. SPECIFICATIONS OF SOME LIKELIHOOD BASED APPROACHES

In this section the basic assumptions and specifications used by the three methods are given. The assumptions for the three methods and some other important methods are summarized in Table 1. The implications of some of these assumptions are discussed in this section, while some are examined by simulation studies in Section 3.

Table 1. Summary of the specifications used for some methods. Explanations in the text.

Specification	Neftci (1982)	Hamilton (1989)	Birchenhall et. al. (1999)	HMLin	SRLin	SRnp
Type of next turn known	Yes	No	No	No	Yes	Yes
Parametric function	$E(X t)$	$E(X t)$	$P(C x)$ logistic	$E(X t)$	$E(X t)$	No.
Equal slopes for the two phases	No	No	-	No	Yes	No
Equal variances for the two phases	No	Yes	-	No	Yes	Yes
Equal slopes over time	Yes	Yes	-	Yes	Yes	No
Classification	$P(C x) > 1 - P(t_A < \tau)$	$P(C x) > 0.5$	$P(C x) > 0.5$ and $P(C x) > \hat{p}$	$P(C X) > 0.5$	$P(C X) > k_{SRLin}; MRL^0 \text{ fix}$	$P(C X) > k_{SRnp}; MRL^0 \text{ fix}$
“Informative” distribution of transitions	Yes	Yes	-	Yes	No	No
Time dependent transition probability	Yes	No	-	No	No	No
Auto Correlation	No	Yes	No	No	No	No

By monitoring the movements of a leading economic indicator, we have an instrument for predicting the turning points of the general business cycle. The aim is to detect a change from expansion to recession (or vice versa) in the leading indicator as soon as possible after it has occurred. Some of the likelihood based methods use an HMM, to

describe the underlying process that changes at an unknown time. An additional aim when using HMM in addition to detecting the change from one phase to another often is to determine a whole chain of phases. That additional aim is not treated in this paper. Only detection of the last change of phases is considered. Thus, the vocabulary of statistical surveillance is suitable.

2.1 Model within each expansion- and recession state

Denote the process under observation by X and the observations available at time t by X_t : $X(1), X(2), \dots, X(t)$. Time $t=1$ is the first time point in a period of special interest. Here we discuss the detection of a peak in the expected value of X : $\mu(1), \mu(2), \dots$ but the problem is equivalent for the detection of a trough. For a peak we have

$$\begin{cases} \mu(1) \leq \mu(2) \leq \dots \leq \mu(t), & t < \tau \\ \mu(1) \leq \dots \leq \mu(\tau-1) \text{ and } \mu(\tau-1) \geq \dots \geq \mu(t), & t \geq \tau \end{cases} \quad (1)$$

where $t=1$ is in a period of expansion and τ is the time of the turn. Note that $\mu(t)$ is monotonic within each state. Observe that the dependency of $\mu(t)$ on τ makes $\mu(t)$ a stochastic variable.

The SRnp method uses only the monotonicity restriction, and is not based on any other assumptions of the shape of the regression.

Many HMM approaches relies on constant differences $\mu(t) - \mu(t-1)$ within each phase. This agrees with the linearity within each phase used for the HMLin method. That is

$$\mu(t) = \begin{cases} \beta_0 + \beta_1 \cdot t, & t < \tau \\ \beta_0 + \beta_1 \cdot (\tau - 1) - \beta_2 \cdot (t - \tau + 1), & t \geq \tau \end{cases}$$

The SRLin method is also based on the assumption of linearity within phases and in addition on the assumption that the slopes for the two phases are equal in absolute values. For research on asymmetry of the business cycle, see Neftci (1984), Falk (1986), McQueen and Thorley (1993). The effect of a non-symmetric turning point on the performance of the SRnp method is studied in Andersson (2001). The effect of misspecification of the SRLin method is examined by the Monte Carlo study in Section 3.

If there is evidence of considerable heteroscedasticity, then the observations should have different weights in the alarm statistic. Sometimes, as here, the logarithm transformation is used for variance stabilization and equal weights are used.

The model discussed here is:

$$X(t) = \mu(t) + \varepsilon(t), \quad (2)$$

where $\varepsilon(t) \sim \text{iid } N[0; \sigma^2]$. The assumptions in (2) might be too simple for some applications, but are used here to emphasize the basic inferential issues. These assumptions are also the ones, which most suggested methods are based on. Ways to handle models with seasonal effects, autoregression and multivariate situations are discussed in the related paper by Andersson et al. (2002).

2.2 Event to be detected

For the SRnp method we have the following situation. At decision time s an alarm statistic is used to discriminate between $D(s) = \{\tau > s\}$ and $C(s) = \{\tau \leq s\}$, where τ is the unknown time when the underlying process μ changes from expansion to recession. Knowledge of whether the next turn will be a trough or a peak is assumed. The solutions

for peak- and trough-detection are equivalent, as everything is symmetrical. It is the knowledge per se which is important. For peak detection, i.e. detection of transition from expansion to recession, the SRnp method discriminates between the following two events:

$$\begin{aligned} D_{SRnp}(s): \mu(1) \leq \dots \leq \mu(s) \\ C_{SRnp}(s): \mu(1) \leq \dots \leq \mu(\tau-1) \text{ and } \mu(\tau-1) \geq \mu(\tau) \geq \dots \geq \mu(s) \end{aligned} \quad (3)$$

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

For the SRLin method the aim is to discriminate between D and C , such that

$$\begin{aligned} D_{SRLin}(s): \mu(s) = \beta_0 + \beta_1 \cdot s \\ C_{SRLin}(s) = \{\cup C(\tau)\}, \end{aligned} \quad (4)$$

where $C(\tau): \mu(s) = \beta_0 + \beta_1 \cdot (\tau-1) - \beta_1 \cdot (s-\tau+1)$ and where $\tau = \{1, 2, \dots, s\}$ and β_0 and β_1 are known constants.

For the HMLin method, the situation is such that at decision time s an alarm statistic is used to discriminate between

$$\begin{aligned} D_{HMLin}(s): \mu(s-1) \leq \mu(s), \\ C_{HMLin}(s): \mu(s-1) > \mu(s). \end{aligned} \quad (5)$$

The difference between the events for SRLin and SRnp is only the assumptions on $\mu(t)$. However, for the HMLin method the events are different also in another aspect. The apparently simpler event in the HMLin approach is combined with a more complicated situation regarding the information of previous states. Knowledge of previous states is not utilized in the HMLin expression for the posterior probability. However, both the events $D_{HMLin}(s)$ and $C_{HMLin}(s)$ correspond to families of histories of states. Due to the Markov dependence the probabilities for the history on those earlier states will have an effect. Thus, $C_{SRnp}(s)$ and $C_{SRLin}(s)$ only concern the last turning point, whereas $C_{HMLin}(s)$ includes a family of series of turning points. The effect on knowledge of the type of the next turn will be further examined in Section 2.6.

2.3 Transition probabilities

The probability of a transition from recession to expansion (or vice versa) can be assumed to be time dependent as by e.g. Neftci (1982), Diebold et al. (1994) and Filardo (1994). Even though there are no principal difficulties with this, constant transition probabilities have been used in this paper like in most approaches in the HMM framework, see e.g. Layton (1996). Thus, a geometric distribution for τ is used.

For the SRnp and SRLin method the type of the next turn (peak or trough) is assumed to be known. This is also assumed by e.g. Neftci (1982). Motivations for this assumption are discussed in Section 2.6. It is then sufficient with one transition probability

$$\nu = P(C(t)|D(t-1)) = P(\tau = t | \tau \geq t).$$

This transition probability is the intensity parameter in the geometric distribution for τ . For the HMLin method two transition probabilities are needed since both transition from recession to expansion and vice versa have to be considered in the alarm statistic.

The transition probability from one phase to another is estimated from an earlier period and treated as known constants by the HMLin method. When the posterior probability for HMLin is calculated in the simulation study the maximum likelihood estimates $p_{12} = 0.13$ and $p_{21} = 0.10$ are used. The transition probabilities will have an effect on the weight that different observations will have in the test statistic. The effect of weighting the observations can be assumed to have a minor influence as long as the

estimated transition probabilities are fairly close to the true values. Greater influence can be expected on the alarm rate, since the alarm limit depends on the distribution of τ . By (7) in Section 2.4 we can see that if we have a constant alarm limit 0.5 for the posterior probability, then the alarm limit for the likelihood ratio will depend on $P(D)/P(C)$, which in turn depends on the transition probabilities. Thus, the HMLin method is sensitive to the values of the transition probabilities.

A non-informative prior is used by the SR and SRnp approaches and therefore no estimate of the transition probability is necessary. When the distribution of τ is unknown, this approach is not optimal, but the approximation seems to work well, (see Frisé and Wessman (1999) where the Shiryaev-Roberts approximation is used to detect a change from an in-control level to an out-of-control level). If a prior based on observed data, for example with a high weight for a turning point after 10 quarters, was used, the influence of data would be reduced and the probability of an alarm after 10 quarters would be very high. This balance between experience from earlier periods and data for the present one has to be judged by the users.

The approach by Birchenhall et al. (1999) is similar to that of Hidden Markov Models and the LR approach for surveillance in the respect that it is based on Bayes theorem and likelihood and that it models the probability of the type of regime. However, a major difference is that the classification into different regimes is based on explanatory variables but not on transition from the earlier state.

2.4 Alarm statistics

In all methods discussed here the alarm statistic is based on the likelihood ratio. The full likelihood ratio method (LR) has several optimal properties, see Frisé and de Maré (1991). The expected utility, based on very general functions of the gain of an alarm and the loss of a false alarm, is maximized. One interesting consequence is that the expected delay of an alarm is minimized for a fixed probability of false alarm. The LR method gives an alarm for the first time s for which

$$\text{LR}(s) = \frac{f(x_s|C)}{f(x_s|D)} > k_s, \quad (6)$$

where f is the likelihood function and $k_s = k/(1-k) \cdot (P(D(s))/P(C(s)))$.

It is shown, by Frisé and de Maré (1991), that the posterior probability approach is equivalent to the likelihood ratio approach for the situation where $P(C) = 1 - P(D)$, i.e.

$$\{x_s : P(C|x_s) \geq k\} = \left\{ x_s : \frac{f(x_s|C)}{f(x_s|D)} \geq \frac{P(D) \cdot k}{P(C) \cdot (1-k)} \right\} \quad (7)$$

where k is the alarm limit for the posterior probability. The choice of k is discussed in Section 2.5 on control of false alarms.

2.4.1 HMLin

The posterior probability is utilized by many authors, e.g. Neftci (1982), Hamilton (1989), Lahiri and Wang (1994) and Kim and Nelson (1998). For the HMLin method, the alarm statistic at time s is the posterior probability based on the observations available at time s , $P(C|X_s)$. The time of the alarm, t_A , for the HMLin method is defined as

$$t_A = \min [s: P(C|X_s) > 0.5] \quad (8)$$

and is thus based on all observations. In some papers, for example Koskinen and Öller (1998), the computational formula for a HMM alarm statistic is presented as

$$\frac{P(C(t)|X(t-1) = x(t-1)) \cdot f(x(t)|C(t))}{f(x(t)|x(t-1))},$$

which equals

$$P(C(t)|X(t) = x(t), X(t-1) = x(t-1)),$$

which gives the impression that only the last two observations are included. However, at decision time s , the formula above is used recursively until we have the alarm statistic

$$P(C(s)|X(s) = x(s), X(s-1) = x(s-1), \dots, X(1) = x(1)) = P(C(s)|x_s).$$

If it is assumed that more than one change can occur in the time interval $\{1, s\}$, the recursive formula, used by the HMLin method, is suitable. However, if only the first change during the monitoring period is of interest and it is known a priori if the next turn will be a peak or a trough, then it is advantageous to utilize this information e.g. by the SRLin and SRnp methods.

2.4.2 SRLin

We derive this method from the LR method (6) which has several optimality properties.

Here we have

$$D(s): \mu(s) = \beta_0 + \beta_1 \cdot s$$

$$C(s) = \{\cup C(\tau)\},$$

$$C(\tau): \mu(s) = \beta_0 + \beta_1 \cdot (\tau - 1) - \delta_1 \cdot (s - \tau + 1),$$

where β_0 , β_1 and δ_1 are known constants.

Under the assumption of a normal distribution, the optimal alarm rule LRlin for discriminating between $D(s)$ and $C(s)$ above can be derived to give an alarm for the first time s where

$$\text{LRlin}(s) > k_s,$$

where $\text{LRlin}(s) =$

$$\sum_{j=1}^s v_j \cdot \exp \left[\left(\frac{1}{2\sigma^2} \right) \left(2(-\delta_1 - \beta_1) \sum_{u=j}^s (x(u) \cdot u) + 4\delta_1 \cdot \sum_{u=j}^s (x(u) \cdot (j-1)) + w_j \right) \right]$$

$$\text{with } w_j = (\beta_1^2 - \delta_1^2) \cdot \sum_{u=j}^s u^2 + 4\delta_1^2 \cdot (j-1) \cdot \left(\sum_{u=j}^s (u-j+1) \right) + 2\beta_0 \cdot (\beta_1 + \delta_1) \cdot \sum_{u=j}^s u - 4\beta_0\delta_1 \cdot \sum_{u=j}^s (j-1)$$

$$\text{and } v_j = \frac{P(\tau = j)}{P(\tau \leq s)}, k_s = k/(1-k) \cdot (P(D_{\text{SRLin}}(s))/P(C_{\text{SRLin}}(s))), \text{ where } k \text{ is a constant.}$$

The LRlin(s) statistic is a function of the transition probability $\nu = P(\tau = t | \tau \geq t)$. The Shiryaev-Roberts, SR, approach by Shiryaev (1963) and Roberts (1966) avoids a choice of this value by using equal values of $P(\tau = t)$ for all t . This approach can be motivated either by the limiting distribution when ν tends to zero or by a non-informative prior for τ . The Shiryaev-Roberts approach implies equal weights for the partial likelihood ratios and a constant alarm limit.

The Shiryaev-Roberts method for detection of a symmetric turning point with linear functions as in (4), has the statistic

$$\text{SRlin}(s) = \sum_{j=1}^s \exp \left[\left(\frac{1}{2\sigma^2} \right) \left(4\beta_1 \cdot \sum_{u=j}^s (x(u) \cdot (j-1-u)) + w_j \right) \right]$$

where $w_j = (4 \cdot \beta_1^2 \cdot (j-1) + 4 \cdot \beta_0 \cdot \beta_1) \cdot \sum_{u=j}^s (u-j+1)$. The time of the alarm, t_A , for the SRlin method is

$$t_A = \min [t: \text{SRlin}(t) > k_{\text{SRlin}}] \quad (9)$$

where k_{SRlin} is a constant alarm limit to be determined to satisfy a false alarm property.

2.4.3 SRnp

The nonparametric SRnp method is a Shiryaev-Roberts variant of the maximum likelihood ratio statistic

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

with maximum over the class of all monotonic (D) or unimodal (C) functions respectively.

To detect the change in monotonicity (3) we need the maximum likelihood estimators under the restrictions D_{SRnp} (monotonically increasing) and C_{SRnp} (a turn).

The denominator of MLR(s) is

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

where $\hat{\mu}^D$ is the estimated parameter vector which corresponds to

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

where F^D is the family of μ -vectors such that $\mu(1) \leq \mu(2) \leq \dots \leq \mu(s)$. This means that $\hat{\mu}^D$ is the maximum likelihood estimator of μ under the monotonicity restriction D . This estimator is described by e.g. Robertson et al. (1988).

For the event $C = C_{\text{SRnp}}$ we have $C = \{C_1, C_2, \dots, C_s\}$, where C_j implies

$$\{\mu(1) \leq \dots \leq \mu(j-1), \mu(j-1) \geq \mu(j) \geq \dots\}, \quad j \in \{1, 2, \dots, s\}.$$

In the numerator of MLR(s) we have

$$\begin{aligned} \max f\{x_s | C\} &= \\ & \sum_{j=1}^s \left(\frac{P(\tau = j)}{P(\tau \leq s)} \right) \cdot \left(\max f\{x_s | C_j\} \right) = \\ & \sum_{j=1}^s \left(\frac{P(\tau = j)}{P(\tau \leq s)} \right) \cdot \left(f\{x_s | \hat{\mu}^{C_j}\} \right), \end{aligned}$$

where $\hat{\mu}^{C_j}$ is the estimated parameter vector which corresponds to

$$\max_{\mu \in F^{C_j}} f(x_s | \mu),$$

where F^{C_j} is the family of μ -vectors such that $\mu(1) \leq \dots \leq \mu(j-1)$ and $\mu(j-1) \geq \mu(j) \geq \dots$, where $j = \{1, 2, \dots, s\}$ and where at least one inequality is strict in the second part.

This means that $\hat{\mu}^{C_j}$, $j \in \{1, 2, \dots, s\}$, is the maximum likelihood estimator of μ under the monotonicity restriction C_j . This estimator was given by Frisén (1986).

Thus, μ is estimated using a non-parametric method and the maximum likelihood ratio at decision time s is

$$\text{MLR}(s) = \sum_{j=1}^s \frac{P(\tau = j) f(x_s; \hat{\mu}^{C_j})}{P(\tau \leq s) f(x_s; \hat{\mu}^D)}.$$

The MLR statistic is a function of the distribution of τ . This is avoided by using the Shiryaev-Roberts approach, as described in Section 2.4.2. The alarm statistic of the SRnp method is

$$\text{SRnp}(s) = \sum_{j=1}^s \frac{f(x_s; \hat{\mu}^{C_j})}{f(x_s; \hat{\mu}^D)}.$$

The time for alarm is

$$t_A = \min [t: \text{SRnp}(t) > k_{\text{SRnp}}], \quad (10)$$

where k_{SRnp} is a constant. The method was suggested by Frisén (1994) and evaluated by Andersson (2001) and Andersson (2002).

The inference situation is different from that of estimation of the number and locations of structural breaks in series with a fixed number of observations. Examples of the latter approach are Mudambi (1997) who describes a method based on polynomial regression for confirmation of the existence of structural breaks and identification of the number and locations of the breaks, and Delgado and Hidalgo (2000), who propose a method based on kernel estimators for estimating the location and size of breaks in a non-parametric regression model.

2.5 Control of false alarms

The way in which false alarms for turns are controlled is important. The constants in the alarm rules of Section 2.4 have to be determined. In the general theory and practice of surveillance, the most common way is to control the ARL^0 , (the Average Run Length to the first alarm if the process does not have any turn). Hawkins (1992), Gan (1993) and Andersson (2002) suggest that the control is made by a statistic similar to the ARL^0 , namely the MRL^0 , which is the median run length $M[t_A | \tau = \infty]$. This has several advantages, such as easier interpretations for the skewed distributions and much shorter computer time for calculations.

The time of the alarm, t_A , for the SRLin and SRnp methods, is the first time for which the alarm statistic exceeds a constant. This constant is determined to yield a fixed value of MRL^0 . In the simulation study below, MRL^0 is chosen to agree with the value achieved by the HMLin method.

A direct Bayesian approach, which is often used, is to control the limit for the posterior probability. The limit 0.5 for the posterior probability is often used when classification is made. This approach is also used for HMLin, as an alarm is given as soon as $P(C|X_s) > 0.5$. Zellner et al. (1991) discuss the limit value of the posterior probability in the context of loss functions. If the loss of a false alarm equals that of a missed alarm, then the expected total loss would be minimal if the limit 0.5 is used for the posterior

probability. However, Birchenhall et al. (1999) describe the limit 0.5 as reflecting lack of prior information. They discuss the use of an estimated prior probability instead of 0.5 and give results for an "uncertain region" where the posterior probability is between these values.

The approach used in much theoretical work e.g. Shiryaev (1963) and Frisé and de Maré (1991) and for which optimality theorems are available, is a control of the probability of false alarm

$$P(t_A < \tau) = \sum_{i=1}^{\infty} P(\tau = i) \cdot P(t_A < i | D). \quad (11)$$

The alarm limit is determined to yield a fixed false alarm probability. Neftci (1982) and Lahiri and Wang (1994) use this criterion for alarms for turning points of business cycles.

Chu et al. (1996) advocate monitoring methods for structural change, which have a fixed (asymptotic) probability of any false alarm during an infinite long surveillance period without change. For some applications, this might be important because a strict significance test is in fact the goal. In that case, ordinary statements for hypotheses testing can be made. However, the price for this is high. The expected delay of the detection will be very large (see Pollak and Siegmund (1975)).

2.6 Knowledge of the type of the next turn

In many applications knowledge of whether the next turn will be a peak or a trough is available. This is certainly so for the natural family planning mentioned in the introduction. For business cycles, the confirmation that a time point is a turning point cannot be made directly. Layton and Katsuura (2001) point out that this is a problem for methods which assume knowledge on the type of past regimes for the estimation of parameters. It might be reasonable to think that the confirmation can come after 3 or 4 quarters. Therefore, in the present simulation study, the evaluation period ($t=1$) starts 4 time points after the last turning point of the estimation period.

Knowledge of whether the next turn will be a peak or a trough makes it possible to use only data during the evaluation period. The likelihood ratio statistic of the surveillance approach can then be used. In fact, for the SRnp approach, nothing will be gained by including earlier time points in the analyses. However, without the information on the type of turn, the last observations contain little information and it is important to utilize also information from earlier times. This is a major difference between the HMM approach on one hand and the surveillance approach on the other. The former approach is used for the HMLin method and by e.g. Koskinen and Öller (1998). The latter approach is used for the SRLin and SRnp methods and by e.g. Neftci (1982) and Diebold and Rudebusch (1989). By comparisons of the differences between the complete methods and the differences induced by different specific effects above, we conclude that the knowledge of the type of the next turn is important information.

If information about the type of the next turn is used in the surveillance, it means that the surveillance can be designed for detecting that particular type of turn. Instead of trying to detect both peaks and troughs, the method is designed for just detecting a peak, thereby simplifying the surveillance situation and improving the detecting ability.

When the type of the next turn is known, the events D and C to be discriminated between are identical for the surveillance methods (e.g. SRLin) and the HMM methods (e.g. HMLin) if the same assumptions are made about the other features such as the shape of the regression. It is demonstrated by Frisé and de Maré (1991), that the likelihood

ratio method and the posterior probability approach give the same result as soon as the events D and C are the same and D is the complement to C . Thus, for a known type of the next turn, the HMM approach is identical to the surveillance by the likelihood ratio method.

Past periods with known regime characteristics carry valuable information. Several authors utilize such information for estimation purposes. If the regimes for all past time points are completely known, then an optimal alarm statistic can be based on only the last observation.

3. MONTE CARLO STUDY ON THE EFFECTS OF DIFFERENT SPECIFICATIONS

3.1 Models used for the simulations

The investigation of the effects of different specifications is made for the detection of a turn in an evaluation period with one turn. In the Monte Carlo study the comparisons are made for a situation similar to that of the Swedish industrial production (IP), after seasonal adjustment. For a description of IP, see Öller and Tallbom (1994). The seasonal adjustment is made using regression on seasonal dummies. The time series is illustrated in Figure 1. The data for the period 1970Q1 to 1987Q1 is used in the estimation process for the HMLin method and the period 1987Q2 to 1992Q2 is used for modeling the evaluation period. The expansions and recessions are dated using the records of the National Institute of Economic Research (1992).

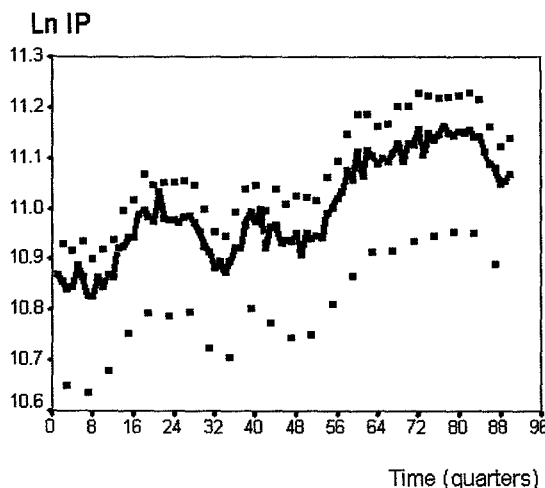


Figure 1: *Industrial production, quarterly data (1970Q1: 1992Q2). The evaluation period starts at 1987Q2, marked with a dashed vertical line. The seasonally adjusted values are connected, while the original values are unconnected.*

3.1.1 Model for event D (no turn)

In order to evaluate the false alarm properties, the event D (no turn) has to be specified. In this case, we need a model of expansion for the whole evaluation period. A linear function was fitted to the expansion phase of the evaluation period (1987Q2: 1989Q3). The observations on X , under event D , are simulated using the same variance as for event C . The model used is

$$X^D(t) = \mu^D(t) + \varepsilon(t), \quad (12)$$

where $\mu^D(t) = 11.194 + 0.0069 \cdot t$ and $\varepsilon(t) \sim \text{iid } N[0; 0.016]$.

3.1.2 Model for event C (a turn)

The aim is to find a model which mimics the actual behavior of the turning point in the evaluation period (1987Q2:1992Q2). The seasonal effects are included as seasonal dummies when the parameters of the regression curve and the standard deviation are estimated. The following model is found to fit well:

$$X^C(t) = \mu^C(t) + \varepsilon(t), \quad (13)$$

$$\text{where } \mu^C(t) = \begin{cases} 11.194 + 0.0066 \cdot t - 0.00017 \cdot t^2 - 0.000015 \cdot t^3, & 1 \leq t \leq 13 \\ 11.340 - 0.0089 \cdot t, & t \geq 14 \end{cases}$$

and $\varepsilon(t) \sim \text{iid } N[0; 0.016]$.

The turning point time (τ), here the peak time, is defined as the first time for which μ decreases since the start and for this model we have $\tau = 10$. This model is used in some of the simulations where the properties for the rounded curve are illustrated. However, it is not suitable for a study of the effect of different values of τ since the growth of the slope is not constant. The different slopes in different parts of the curve will also have an influence when the value of τ is varied. Thus, for examination of the influence of different values of τ , an approximation of the rounded smooth curve is used, where the slopes are constant and equal before and after the peak.

$$X^{C\tau}(t) = \mu^{C\tau}(t) + \varepsilon(t), \quad (14)$$

$$\text{where } \mu^{C\tau}(t) = 11.194 + 0.0069 \cdot t - 2D_1 \cdot 0.0069 \cdot (t - \tau + 1), \quad t = \{1, 2, \dots\},$$

$$\text{and } D_1 = \begin{cases} 1, & t \geq \tau \\ 0, & \text{otherwise} \end{cases}$$

and $\varepsilon(t) \sim \text{iid } N[0; 0.016]$.

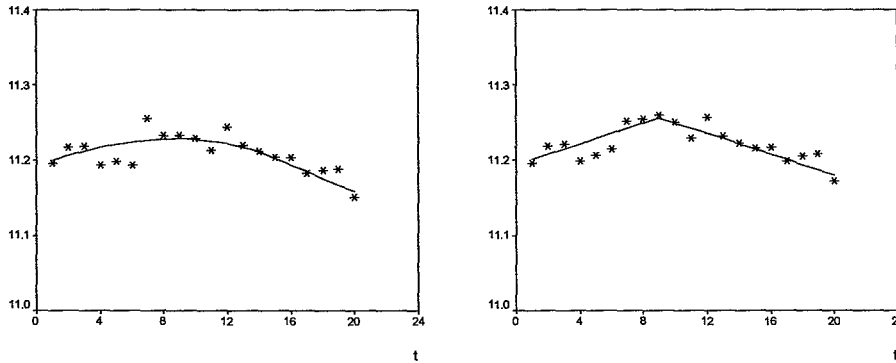


Figure 2: The rounded regression curve and the piecewise linear one with a turn at $t=10$, and one realization for each of the models (13) and (14).

3.1.3 Specifications for the estimation period

Observations not only in the evaluation period, but also in previous expansions and recessions are used by most methods (see e.g. Neftci (1982), Hamilton (1989), Lahiri and Wang (1994)) and here also by the HMLin method. One object in our simulation study is to study the effect of estimation of parameters in the regression function.

Regression curves that are similar to those of the estimation period are determined by fitting one regression model, including seasonal dummies, to each expansion- and recession phase respectively. The intercepts of the regression models are adjusted to

avoid jumps. The resulting chain of polynomials, without the seasonal components, and with the estimated standard deviation, is used as a model for the simulations.

3.2. Control of false alarms

For the Hmlin approach, the alarm limit is the threshold probability 0.5 for the posterior probability. For the expansion situation, D , when there is no turn, the result from the simulation study is that this alarm limit will result in a median run length $MRL^0 = 17$ with a standard error of 0.13.

The alarm limits of the SRnp and SRLin methods are determined by an iterative procedure to yield the same MRL^0 , 17. The standard error of the last estimate of MRL^0 is 0.11 for SRnp and 0.12 for SRLin.

3.3. Evaluation of the effect of different specifications

The evaluations and comparisons of the methods is made using the probability of false alarm, the expected delay of an alarm, the probability of successful detection and the predictive value, as suggested by Frisén (1992).

3.3.1 Comparison between methods with correct specification

Two of the methods, SRLin and Hmlin, assume knowledge about the shape of the regression. The comparisons are first made for the case when the actual function is linear (which is assumed by SRLin and Hmlin) according to (14).

3.3.1.1 False alarm

As is seen in Figure 3, the Hmlin method has more frequent false alarms at early time points, but low alarm probability later compared with that of the SRLin and especially SRnp. The curve for the SRLin method is between the other two. All curves cross at $t=17$. This is due to the construction of comparability: for all three methods the median run length to the first false alarm is set to be $MRL^0 = 17$.

$$P(t_A \leq t)$$

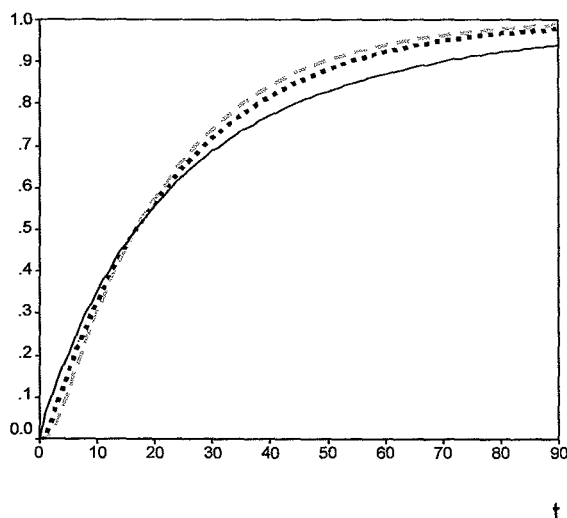


Figure 3: The distribution of the time (quarters) of an alarm conditional on event D (no turn). Hmlin (—), SRnp (---), SRLin (···).

The total probability of a false alarm, $P(t_A < \tau)$ depends on the distribution of τ . When τ has a geometric distribution, the false alarm probability is the smallest for the SRnp, and the largest for the Hmlin. As a result of the assumption of a geometric distribution for τ , the alarms at the beginning have a great influence on $P(t_A < \tau)$. The large false alarm probability for Hmlin is a result of the error-spending curve with many early alarms.

3.3.1.2 Delay of the alarm

To illustrate how the probability of an alarm is changed at the turning point, the run length distributions when $\tau=10$ are given in Figure 4 for the three methods. An immediate alarm at the turning point is desired but a probability of a delay is unavoidable.

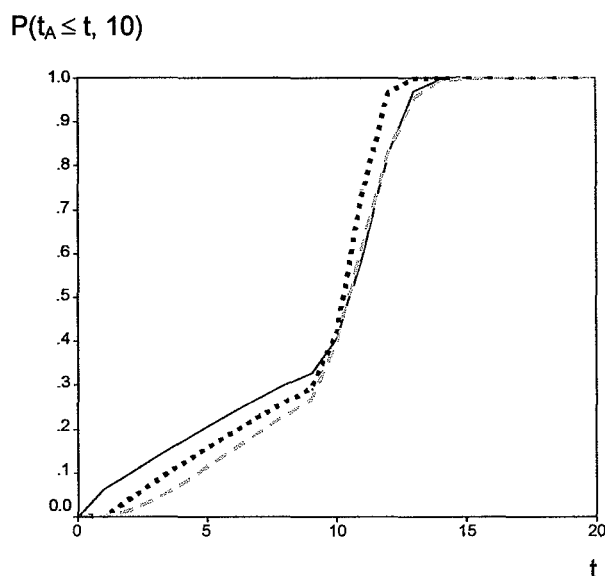


Figure 4: The distribution of the time of an alarm for the piecewise linear curve, $\tau=10$. Hmlin (—), SRnp (---), SRLin (···).

The conditional expected delay time of an alarm, $CED(\tau) = E[t_A - \tau | t_A \geq \tau]$ gives a measure of how well a method works. Also the conditional median delay, $CMD(\tau)$ was calculated but gave similar results for the comparison between methods (but lower values) and these are thus not reported.

The rounded model (13) gave shorter delay than the piecewise linear one (14) (the details are not reported here) for the Hmlin method, while there was no difference for the SRnp method. One reason for this is that the rounded curve deviates from the Hmlin-model already before the turn.

The delay times for the piecewise linear case are summarized by $CED(\tau)$ in Figure 5. For $\tau \leq 20$ we have a standard error less than 0.009.

The conditional expected delay is worse for SRnp than for Hmlin, for small values of τ , $\tau < 4$. After that, the delay is slightly shorter for SRnp, compared to Hmlin. The effect of τ is large for SRnp for small values of τ . However, an asymptote is reached at about $\tau=10$. For Hmlin, the effect of τ is very small. A very slight increase in the conditional

expected delay can be observed as τ increases. The SRlin method has the shortest delay for every τ . Both SRlin and HMLin reach their respective asymptote very early. The reason is that both these methods assume the correct parametric function for the turn (a piecewise linear function). The SRnp method needs more observations in the beginning to have enough evidence of a turn.

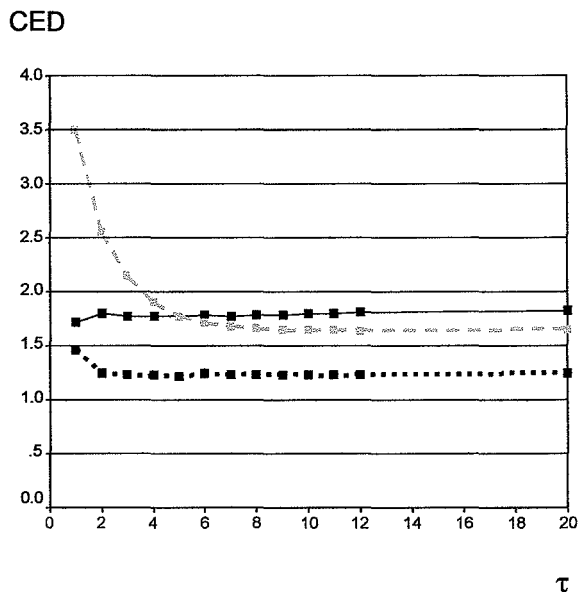


Figure 5: *Conditional expected delay for the piecewise linear curve with turn at τ . HMLin (—■), SRnp (---■), SRlin (···■).*

The conditional expected delay is further summarized under the assumption of a geometric distribution for τ , by

$$ED = \sum_{i=1}^{\infty} CED(\tau) \cdot P(\tau = i).$$

The expected delay, ED, is minimized by the full likelihood ratio method when methods with the same probability of false alarm are compared. Using $\nu=0.10$ in the geometric distribution, ED is 1.79 for HMLin, 1.99 for SRnp and 1.26 for SRlin (with standard errors less than 0.0026). Thus, for correct specification the nonparametric SRnp method is not as good as the parametric ones. However the effect of knowledge of the type of the next turn, as used by the SRlin method is greater than the effect of correct parametric specification.

3.3.1.3 Probability of successful detection

Sometimes an alarm that comes too late is worthless. Thus it might be useful to complement the measure of expected delay with the probability of successful detection within d time points, $PSD = P((t_A - \tau) \leq d | t_A > \tau = \tau_0)$. This measure is given in Figure 6 for $\tau=10$ and the piecewise linear model (14). The number of replicates is large so the standard error is less than 0.003 for each point of the curve.

The PSD curves are very similar for the HMLin and SRnp methods. For SRnp, Andersson (2001) proved that the PSD increases as the post peak slope grows steeper. The effect of the rounded curve, compared to the piecewise linear curve, is twofold: A rounded peak results in an increase in the alarm statistic just before the peak. This means

that only a small increase in the alarm statistic is needed to call an alarm at the time points just after the peak. The result is an increase in the PSD. On the other hand, the characteristics of the peak just after the turning point (rounded or linear) will affect the alarm statistic and the PSD in opposite direction, thus resulting in a decreased PSD for a rounded peak. The SRLin method has the best PSD as could be expected since it utilizes both a known parametric model and a known type of next turn.

PSD(d, 10)

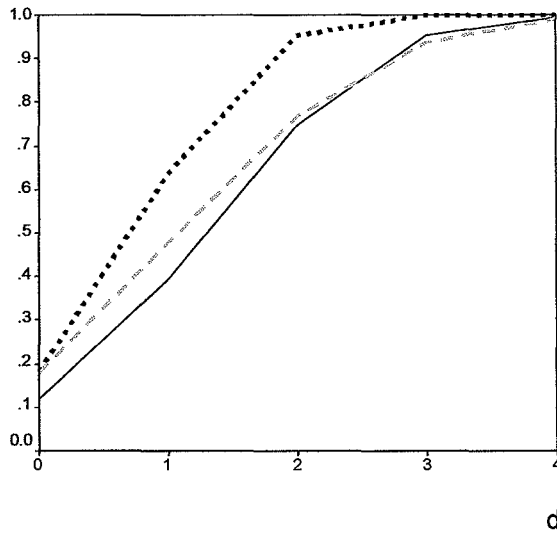


Figure 6: Probability of successful detection within d time points for the piecewise linear curve, $\tau=10$. HMLin (—), SRnp (---), SRLin (···).

3.3.1.4 Predictive value of an alarm

The predictive value of an alarm at time t , $PV(t) = P(\tau \leq t | t_A = t)$ reflects the trust you should have in an alarm. In Figure 7 the predictive value for $t = \{1, 2, \dots, 12\}$ under the assumption of a geometric distribution with intensity $\nu=0.1$ is presented. For $t = 1$ the exact value is calculated and for $t = \{2, \dots, 12\}$, simulated values are used. From Figure 7 it is evident that the price for the high alarm probability in the first point for the HMLin method is that those alarms are of little value. Since the predictive value is only 0.2, an alarm would hardly motivate any action.

Both SRLin and HMLin reach their respective asymptotes early. The development for SRnp is a little different. The predictive value of SRnp increases until $t = 6$, after that the predictive value decreases slightly and reaches the same asymptote as SRLin at approximately $t = 10$. The SRnp method places no parametric restrictions on the turning point curve. All information about the curve comes from data. For small values of t the number of observed data is very small and thus the data have to be very extreme in order to call an alarm. However, as t increases (and the number of observations increases) the information about the curve is improved and at $t = 10$, SRnp has the same predictive value as SRLin.

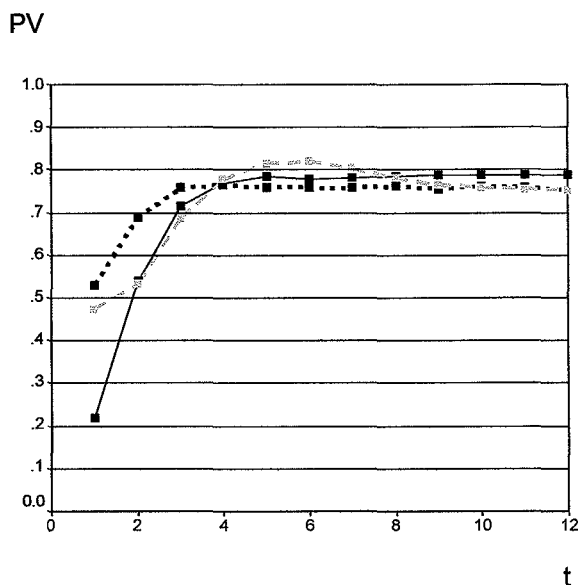


Figure 7: Predictive value of an alarm at time t for $\nu = 0.1$. Hmlin (— ■), SRnp (--- ■), SRLin (··· ■).

3.3.2 Start of the evaluation period

For the methods SRnp and SRLin it does not matter for the false alarm probabilities if the evaluation period is started directly after a regime shift or a little later. However, for the Hmlin method this has a great influence. The reason for this is the difference in the knowledge of the type of the next turn. The probability of classifying the state as a continued recession is very high just after a trough if you do not have the information that the change of regime has already happened. The run length distribution and particularly the probability of a false alarm at the first time point is highly dependent on where the evaluation period begins in relation to the last change. To begin the monitoring at a turning point results in a very large false alarm probability (0.22) at $t=1$ for the Hmlin method while it is reasonable (0.064) when started at $t=4$, as is done here.

3.3.3 Effect of wrongly specified slopes

If the slopes are estimated from a short period, then the parameters might be severely misspecified. If the pattern is not stable, then even a long period for estimation will result in estimates that are not very useful without information about the natural variation of the pattern. The estimation procedures will result in a stochastic deviation from the relevant values. Few methods and neither SRLin nor Hmlin incorporates this uncertainty in the alarm conditions. The effect of a wrong specification of the regression coefficient is investigated for the SRLin method for two situations, namely an expansion and a turn at time τ . Here we allow for unsymmetrical turning points. The examples of misspecification are chosen such that they represent values between which approximately 95% of the expansion estimates would be ($\text{sd}[\hat{\beta}_1] = 0.0009$), with the estimation procedure used at the Swedish National Institute (Koskinen and Öller (1998)).

3.3.3.1 Only the post-peak slope is incorrectly specified

We start with the situation when only the post-peak slope is incorrectly specified. The likelihood ratio

$$LR(s) = \frac{f(x_s|C)}{f(x_s|D)}$$

is stochastically smaller the greater the difference between the events D and C , if the event D is true.

We will now look at the situation in detail. The correct regression is

$$\mu(t) = \begin{cases} \beta_0 + \beta_1 \cdot t, & t < \tau \\ \beta_0 + \beta_1 \cdot (\tau - 1) - \beta_1 \cdot (t - \tau + 1), & t \geq \tau \end{cases}$$

but the specified regression, used in the alarm statistic, is

$$M(t) = \begin{cases} \beta_0 + \beta_1 \cdot t, & t < \tau \\ \beta_0 + \beta_1 \cdot (\tau - 1) - (\beta_1 + \gamma) \cdot (t - \tau + 1), & t \geq \tau. \end{cases}$$

$\beta_1=0.0069$ and $\gamma = \{0.0018, -0.0018\}$. The correct case and the two misspecifications are illustrated in Figure 8.

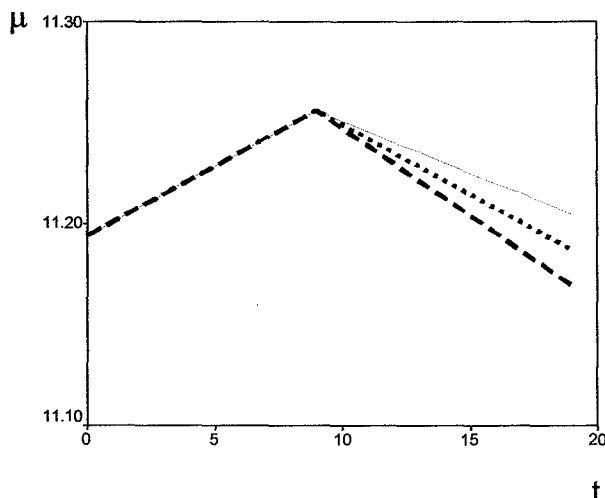


Figure 8: The piecewise linear regression curve for a turn at 10. The post-peak slope is correctly specified (\cdots), too flat (---) and too steep (---).

At decision time 1 we have that the SRLin statistic, using β_1 , is

$$SRLin(1) = \exp\left[\left(\frac{1}{2\sigma^2}\right)(4\beta_0\beta_1 - x(1) \cdot (4\beta_1))\right]$$

whereas the SRLin statistic, using $(\beta_1 + \gamma)$, is

$$S\tilde{R}(1) = \exp\left[\left(\frac{1}{2\sigma^2}\right)(4\beta_0\beta_1 - \gamma^2 - 2\beta_1\gamma + 2\beta_0\gamma - x(1) \cdot (4\beta_1 + 2\gamma))\right].$$

The statistic is decreasing with γ as soon as $X(1) > (\beta_0 - \beta_1)$. This is in agreement with the general result that a likelihood ratio is stochastically smaller for the D event if the difference between the events is large. However, when the alarm limit is adjusted to give the same MRL^0 , the situation changes. For this case (but not the next that will be studied) the adjustment is minor and no details are reported.

There are no dramatic differences for the conditional expected delay, when comparing $\gamma=0$ and $\gamma \neq 0$. For the case of too steep a post-peak slope the CED(1) is slightly smaller than for a correctly specified slope. The deviation is of about the same size but of opposite sign for the case of too flat a post-peak slope. The difference between a correctly and an incorrectly specified slope is small at the beginning of the recession. Since the delays are expected to be small, the misspecification in slope has a minor effect. For that reason it is natural that we do not observe any large differences between SRLin with a correctly specified slope and SRLin with an incorrectly specified slope

The predictive value is very similar between $\gamma=0$ and $\gamma \in \{0.0018, -0.0018\}$, except at $t=1$. The difference is due to the difference in the error-spending curve. For $\gamma = -0.0018$, the alarm statistic is optimized to detect a smaller change (flatter post-peak slope) and therefore the alarms are located later on. The result is few alarms at early time points, which results in a high predictive value for $t=1$. The opposite holds for $\gamma = 0.0018$.

3.3.3.2 Both the pre-peak slope and the post-peak slope are incorrectly specified

Now we look at the situation when both the pre-peak slope and the post-peak slope are incorrectly specified so that the correct regression is

$$\mu(t) = \begin{cases} \beta_0 + \beta_1 \cdot t, & t < \tau \\ \beta_0 + \beta_1 \cdot (\tau - 1) - \beta_1 \cdot (t - \tau + 1), & t \geq \tau \end{cases}$$

but the specified regression, used in the alarm statistic, is

$$M(t) = \begin{cases} \beta_0 + (\beta_1 + \gamma) \cdot t, & t < \tau \\ \beta_0 + (\beta_1 + \gamma) \cdot (\tau - 1) - (\beta_1 + \gamma) \cdot (t - \tau + 1), & t \geq \tau \end{cases}$$

where $\beta_1=0.0069$ and $\gamma = 0.0018$.

This case will suffice to demonstrate the dramatic difference compared with the earlier case with incorrect specification only in C .

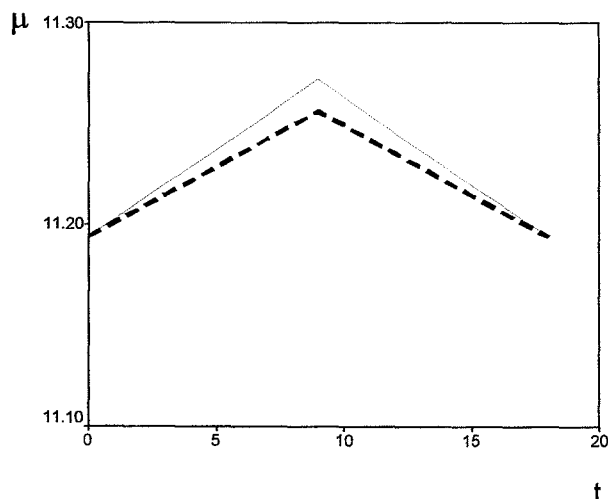


Figure 9: The piecewise linear regression curve for a turn at $t = 10$. Both pre-peak and post-peak slopes are correctly specified (---), too steep (—).

If the same alarm limit that was used in the correctly specified case would be kept, there would be a great increase in false alarms. The probability of a false alarm would, for the misspecified case, increase with time as the difference to the true D -state will

increase. The limit will be determined so that the $MRL^0 = 17$. At that time the difference between the states are enormous. Thus the limit will be changed much to compensate for this (from 7 to 960). The change of the alarm limit will create a completely new situation, as is seen below.

$P(t_A \leq t)$

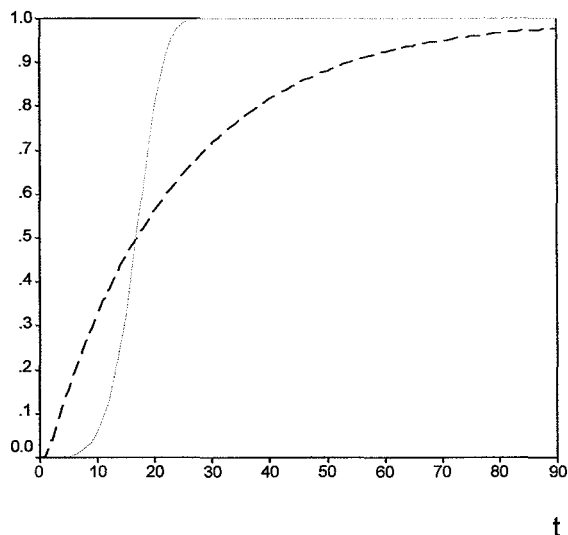


Figure 10: *The distribution of the time of an alarm, conditional on event D (no turn). Both pre-peak and post-peak slopes are correctly specified (---), too steep (—).*

The low false alarm probability for small t for the incorrect specification ($\gamma=0.0018$) in Figure 10 is due to the increasing difference between the true and specified states. For small t the difference is small. However, as t increases, so does the difference. Thus, the likelihood for the specified D -state decreases and therefore, the alarm probability increases.

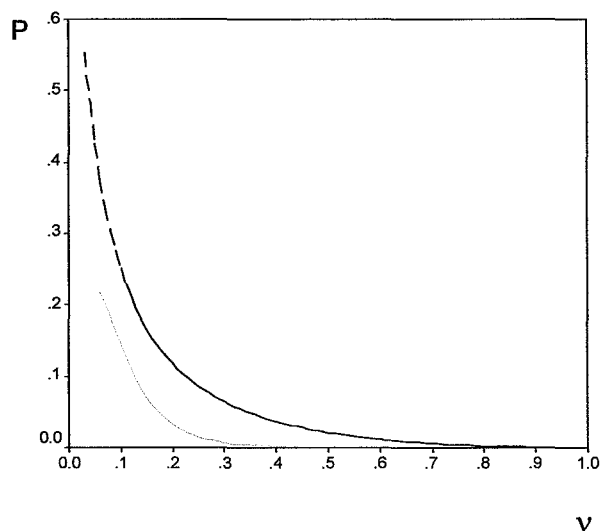


Figure 11: *The probability of a false alarm when τ is geometrically distributed with the parameter v . Both pre-peak and post-peak slopes are correctly specified (---), too steep (—).*

The false alarm probability is the smallest when using the incorrect specified slopes for every value of the intensity, ν . The low false alarm probability is a result of the error-spending curve with few early alarms, in contrast to using correctly specified slopes that result in many early alarms.

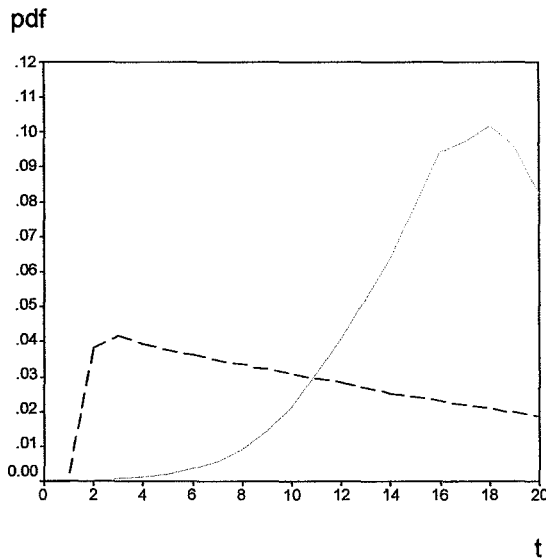


Figure 12: The density of the time of an alarm, conditional on event D (no turn). Both pre-peak and post-peak slopes are correctly specified (---), too steep (—).

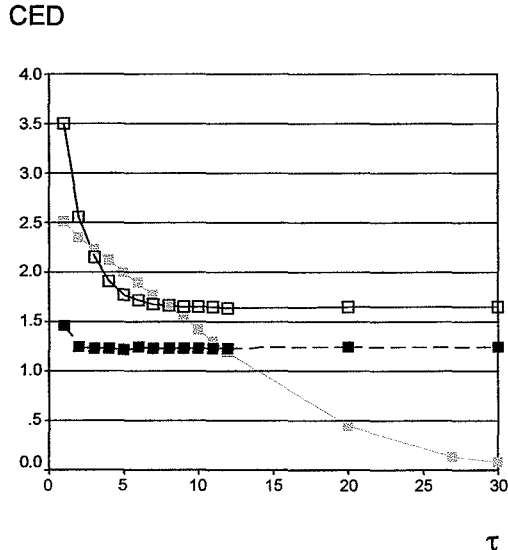


Figure 13: Conditional expected delay for a turn at τ . SRlin correctly specified slopes (--- ■), SRlin too steep slopes (— ■), SRnp (— □).

For small values of τ , the conditional delay is longer if both slopes (expansion and recession) are over-estimated, as seen in Figure 13. As τ increases, the conditional delay decreases towards an asymptote, zero. Thus for the situation where both slopes are mis-specified (too steep) the resulting delay is large if the turning point occurs early. If the turning point occurs late, the CED is zero (all alarms are false alarms). We also see that

although the non-parametric approach, SRnp, has a long delay time for early turns, it quickly reaches a reasonable asymptotic value of CED. We have standard errors less than 0.009.

The predictive value of an alarm, under the assumption of a geometric distribution with intensity ν is shown in Figure 14.

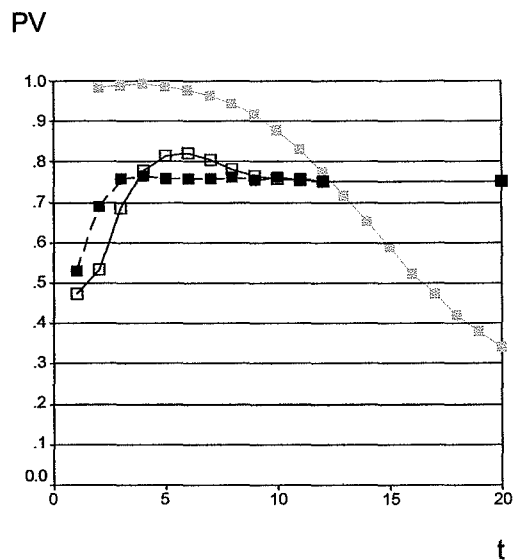


Figure 14: Predictive value of an alarm at time t for $\nu = 0.1$. SRlin correctly specified slopes (--- ■), SRlin too steep slopes (— ■), SRnp(— □).

The predictive value is high for small values of t , as a result of the small false alarm probability. For larger values of t the predictive value decreases. We also see that the PV for SRnp is lower than that of SRlin (correct) for small values of t , but the asymptotic PV (which is the same as for SRlin (correct)) is reached quickly.

The conclusion is that wrong assumptions about slopes may give very bad properties and that the SRnp method gives a safe way to avoid this.

4. DISCUSSION

Similarities between apparently different approaches are demonstrated. This might be a base for combining knowledge from several areas. Different approaches, HMM and statistical surveillance, are expressed in different ways but are equivalent when the assumptions are the same.

The effect of knowledge of the type of the next turn has a major impact on the test statistics if this knowledge is utilized in the likelihood expressions. In practice, there should be no doubts about the type of the next turn as soon as the previous one is verified.

The control of false alarms by the average time to a false alarm or by the probability of a false alarm is commonly used in surveillance. In the comparisons by simulations, all methods are adjusted to have the same median run length, MRL^0 , when no turning point occurs.

The Bayesian requirement that an alarm should be called whenever the posterior probability is greater than 0.5, is commonly used in the literature on Hidden Markov Models. Then, the assumed transition probability has a major impact on the false alarm tendency. Another drawback is that the fixed limit 0.5 might not give the appropriate false alarm rate for all applications. To decrease the false alarm rate, approaches such as smoothing are sometimes used.

As is seen in Figure 3, the HMLin method has frequent false alarms at early time points, but low alarm probability later. This is due to the lack of utilization by the HMM approach of the knowledge of the type of the next turn. This allocation of the alarm probability to the beginning also implies short delay for alarm for early changes, but slightly longer delays for changes that occur late (see Figure 5 and 6). However, the usefulness of these early alarms can be questioned, as the predicted value is very low. The probability that a shift has occurred is only 0.2 when an alarm is made at the first time point. Observe that this is the case in spite of the fact that the posterior probability of a shift was above 0.5. We have thus two different measures of the trust in a shift. Which one of these that is most easily interpreted, by those who make the actual economical decisions, can be discussed.

Most studies in this area assume a constant transition probability, which implies a geometric distribution for the time of the turn. A geometric density has the highest values at the early times. This is not in accordance with reality for business cycles. If evaluated with historical data, we would get the best predictors by using the density that agrees with history. Technically this is easily done by the likelihood ratio methods. However, an important aim is to make an alarm for turn also when this happens at an unexpected time. Thus, here we prefer to use a non-informative prior for the time of the shift in the suggested SRLin and SRnp methods.

Parametric models contain information, which should be used whenever it is reliable. However, it is demonstrated that misspecifications which are probable in practice by the estimation procedure, have a major impact. One advantage of the non-parametric approach is that it works also when reliable information about the parametric function is not available. Also important is that the non-parametric method does not assume that all phases of the same type have the same level and parametric shape. In practice, this varies a lot. The SRnp method only uses the monotonicity change and not the level. The safe way by the SRnp method might be preferred, especially since there is no major loss of efficiency.

Not all differences between the methods have been examined here. Hopefully, the ones that are analyzed will give some insight into the influence of assumptions used in some papers. An important issue for future research is to examine which characteristic of the leading index is the best predictor for a turn in the business cycle. The question remains whether it is the level as in Birchenhall et al. (1999), transition and level as in Hamilton (1989) and Koskinen and Öller (1998) or transition and change in monotonicity as in Frisé (1994) and Andersson (2002) that is most useful. The techniques of multivariate surveillance might be useful for this.

ACKNOWLEDGEMENTS

Constructive comments by Lars-Erik Öller, National Institute of Economic Research, Stockholm, Sweden, are gratefully acknowledged. This work was supported by the Swedish Council for Research in the Humanities and Social Sciences by grant F0473/2000.

REFERENCES

- Andersson E (2001) Turning point detection using non-parametric statistical surveillance. Evaluation of some influential factors. Research Report, Department of Statistics, Göteborg University, Sweden, 2001:3
- Andersson E. 2002. Monitoring cyclical processes - a nonparametric approach. *Journal of Applied Statistics*, **29**.
- Andersson E, Bock D, Frisén M (2002) Some statistical aspects on methods for detection of turning points in business cycles. Research Report, Department of Statistics, Göteborg University, Sweden, 2002:7
- Bidarkota PV. 2001. Alternative regime switching models for forecasting inflation. *Journal of Forecasting*, **20**: 21-35.
- Birchenhall CR, Jessen H, Osborn DR, Simpson P. 1999. Predicting US business-cycle regimes. *Journal of Business & Economic Statistics*, **17**: 313-323.
- Chu C-SJ, Stinchcombe M, White H. 1996. Monitoring structural change. *Econometrica*, **64**: 1045-1065.
- Delgado MA, Hidalgo J. 2000. Nonparametric inference on structural breaks. *Journal of Econometrics*, **96**: 113-144.
- Diebold FX, Lee J-H, Weinbach GC (1994) Regime Switching with Time-varying Transition Probabilities. In *Nonstationary Time Series Analysis* (Ed, Hargreaves C) Oxford University Press, Oxford, pp. 283-302.
- Diebold FX, Rudebusch GD. 1989. Scoring the leading indicators. *Journal of Business*, **62**: 369-91.
- Falk B. 1986. Further Evidence on the Asymmetric Behavior of Economic Time Series over the Business Cycle. *Journal of Political Economy*, **94**: 1097-1109.
- Filardo AJ. 1994. Business-Cycle Phases and Their Transitional Dynamics. *Journal of Business & Economic Statistics*, **12**: 299-308.
- Frisén M. 1986. Unimodal regression. *The Statistician*, **35**: 479-485.
- Frisén M. 1992. Evaluations of Methods for Statistical Surveillance. *Statistics in Medicine*, **11**: 1489-1502.
- Frisén M (1994) Statistical Surveillance of Business Cycles. Research Report, Department of Statistics, Göteborg University, Sweden, 1994:1 (Revised 2000)
- Frisén M (2002) Statistical surveillance. Optimality and methods. Research Report, 2002:2 Department of Statistics, Göteborg University,
- Frisén M, de Maré J. 1991. Optimal Surveillance. *Biometrika*, **78**: 271-80.
- Frisén M, Wessman P. 1999. Evaluations of likelihood ratio methods for surveillance. Differences and robustness. *Communications in Statistics. Simulations and Computations*, **28**: 597-622.
- Gan F. 1993. An optimal-design of EWMA control charts based on median run-length. *Journal of Statistical Computation and Simulation*, **45**: 169-184.
- Gregoir S, Lengart F. 2000. Measuring the Probability of a business Cycle Turning Point by Using a Multivariate Qualitative Hidden Markov Model. *Journal of Forecasting*, **19**: 81-102.
- Hackl P, Westlund AH. 1989. Statistical analysis of "Structural Change". *Empirical Economics*, **14**: 167-192.
- Hamilton JD. 1989. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, **57**: 357-384.
- Hawkins DL. 1992. Detecting shifts in functions of multivariate location and covariance parameters. *Journal of Statistical Planning and Inference*, **33**: 233-244.
- Jun DB, Joo YJ. 1993. Predicting turning points in business cycles by detection of slope changes in the leading composite index. *Journal of Forecasting*, **12**: 197-213.
- Kim C-J, Nelson CR. 1998. Business cycle turning points, a new coincident index, and tests of duration dependence based on a dynamic factor model with regime switching. *Review of Economics and Statistics*, **80**: 188-201.
- Kontolemis ZG. 2001. Analysis of the US business cycle with a vector-Markov- switching model. *Journal of Forecasting*, **20**: 47-61.
- Koskinen L, Öller L-E (1998) A Classifying Procedure for Signalling Turning Points. Research Report, National institute of economic research. Konjunkturinstitutet, Ekonomiska rådet, Stockholm, Sweden, 59
- Lahiri K, Moore G. 1991. *Leading economic indicators: New approaches and forecasting record*, Cambridge University Press: Cambridge
- Lahiri K, Wang JG. 1994. Predicting cyclical turning points with a leading index in a Markov Switching model. *Journal of Forecasting*, **13**: 245-263.

- Lai TL. 1995. Sequential Change-point Detection in Quality-Control and Dynamical-Systems. *Journal of the Royal Statistical Society B*, **57**: 613-658.
- Layton AP. 1996. Dating and predicting phase changes in the U.S. business cycle. *International Journal of Forecasting*, **12**: 417-428.
- Layton AP, Katsuura M. 2001. Comparison of regime switching, probit and logit models in dating and forecasting US business cycles. *International Journal of Forecasting*, **17**: 403-417.
- Li DT, Dorfman JH. 1996. Predicting turning points through the integration of multiple models. *Journal of Business and Economic Statistics*, **14**: 421-428.
- McQueen G, Thorley S. 1993. Asymmetric business cycle turning points. *Journal of Monetary Economics*, **31**: 341-362.
- Mudambi R. 1997. Estimating turning points using polynomial regression. *Journal of Applied Statistics*, **24**: 723-732.
- National Institute of Economic Research (1992) Konjunkturläget Maj 1992. Research Report, National Institute of Economic Research, Stockholm, Sweden,
- Neftci S. 1982. Optimal prediction of cyclical downturns. *Journal of Economic Dynamics and Control*, **4**: 225-41.
- Neftci S. 1984. Are Economic Time Series Asymmetric over the Business Cycle? *Journal of Political Economy*, **92**: 307-328.
- Pollak M, Siegmund D. 1975. Approximations to the Expected Sample Size of Certain Sequential Tests. *Annals of Statistics*, **3**: 1267-1282.
- Roberts SW. 1966. A Comparison of some Control Chart Procedures. *Technometrics*, **8**: 411-430.
- Robertson T, Wright FT, Dykstra RL. 1988. *Order Restricted Statistical Inference*, John Wiley & Sons Ltd.
- Shiryayev AN. 1963. On optimum methods in quickest detection problems. *Theory of Probability and its Applications*, **8**: 22-46.
- Srivastava MS, Wu Y. 1993. Comparison of EWMA, CUSUM and Shiryayev-Roberts Procedures for Detecting a Shift in the Mean. *Annals of Statistics*, **21**: 645-670.
- Westlund AH, Zackrisson U. 1986. On the prediction of structurally varying systems. *Technological Forecasting and Social Change*, **30**: 63-72.
- Wetherhill GB, Brown DW. 1991. *Statistical Process Control: Theory and Practice*, 1. Chapman and Hill: London.
- Zarnowitz V, Moore GH. 1982. Sequential signals of recessions and recovery. *Journal of Business*, **55**: 57-85.
- Zellner A, Hong C, Min C-K. 1991. Forecasting Turning Points in International Output Growth Rates Using Bayesian Exponentially Weighted Autoregression, Time-Varying Parameter, and Pooling Techniques. *Journal of Econometrics*, **49**: 275-304.

Author's biographies:

Eva Andersson is postdoc research associate in Statistics at Göteborg University. Her research interests include business cycles and statistical surveillance.

David Bock is graduate student in Statistics at Göteborg University. His research interests include financial statistics and statistical surveillance.

Marianne Frisén is professor of Statistics at Göteborg University. Her research interests include statistical surveillance and nonparametric methods.

Authors addresses':

Eva Andersson, David Bock and Marianne Frisén, Department of Statistics, Göteborg University, Box 660, SE 40530 Göteborg, Sweden.

Research Report

- | | | |
|--------|--|--|
| 2001:1 | Holgersson, H.E.T.: | On assessing multivariate normality. |
| 2001:2 | Sonesson, C. & Bock, D.: | Statistical issues in public health monitoring – A review and discussion. |
| 2001:3 | Andersson, E.: | Turning point detection using non-parametric statistical surveillance. Evaluation of some influential factors. |
| 2001:4 | Andersson, E. & Bock, D.: | On seasonal filters and monotonicity. |
| 2001:5 | Andersson, E., Bock, D. & Friséén, M.: | Likelihood based methods for detection of turning points in business cycles. A comparative study. |
| 2001:6 | Sonesson, C.: | Evaluations of some exponentially weighted moving average methods. |
| 2001:7 | Sonesson, C.: | Statistical surveillance. Exponentially weighted moving average methods and public health monitoring. |
| 2002:1 | Friséén, M. & Sonesson, C.: | Optimal surveillance based on exponentially weighted moving averages. |
| 2002:2 | Friséén, M.: | Statistical surveillance. Optimality and methods. |
| 2002:3 | Jonsson, R. & Persson, A.: | Bayes prediction of binary outcomes based on correlated discrete predictors. |
| 2002:4 | Persson, A.: | Prediction of work resumption among men and women with lower back- and neck pain in a Swedish population. |
| 2002:5 | Persson, A.: | Prediction of work resumption in theory and practice. |
| 2002:6 | Andersson, E., Bock, D. & Friséén, M.: | Detection of turning points in business cycles. |
| 2002:7 | Andersson, E., Bock, D. & Friséén, M.: | Some statistical aspects on methods for detection of turning points in business cycles. |