



Research Report
Department of Statistics
Göteborg University
Sweden

**Likelihood based methods for
detection of turning points
in business cycles**

A comparative study

**Eva Andersson
David Bock
Marianne Frisé**

**Research Report 2001:5
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			

LIKELIHOOD BASED METHODS FOR DETECTION OF TURNING POINTS IN BUSINESS CYCLES

A comparative study

E. Andersson
D. Bock
M. Frisé

Department of Statistics
Göteborg University
SE 405 30 Göteborg, Sweden

ABSTRACT

Methods for on-line monitoring of business cycles are compared with respect to the ability of early prediction of the next turn by an alarm for a turn in a leading index. Three likelihood-based methods for turning point detection are compared in detail by using the theory of statistical surveillance and by simulations. One of the methods is based on a Hidden Markov Model. Another includes a non-parametric estimation procedure. Evaluations are made of several features such as knowledge of shape and parameters of the curve, types and probabilities of transitions and smoothing. 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. The three methods are also used to analyze an actual data set of a period of the Swedish industrial production. The relative merits of evaluation of methods by one real data set or by simulations are discussed.

Key words: Business cycles; Early warning; Monitoring; Optimal; Likelihood ratio; Bayes; Markov; HMM; Switching regime; Turning point; Non-parametric.

1 INTRODUCTION

Timely prediction of a turn in the business cycle is important for both government and industry. The turn is a change from a phase of recession to one of expansion (or vice versa). The tardiness of stabilization policies induces unintended effects (see Öller (1987)). Predictions of the turning points 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 signals 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), Zellner et al. (1991), Jun and Joo

(1993), Lahiri and Wang (1994), Li and Dorfman (1996) and Birchenhall et al. (1999).

There are two distinct but related approaches to the characterization and dating of the business cycle, as pointed out by e.g. Diebold and Rudebusch (1996), Kim and Nelson (1998) and Birchenhall et al. (1999). One emphasizes the regime shift and the other the common movements of several variables. In this paper the regime approach is pursued, as also in the works by Neftci (1982), Diebold and Rudebusch (1989), Hamilton (1989), Jun and Joo (1993), Lahiri and Wang (1994), Koskinen and Öller (1998) and Birchenhall et al. (1999). The common movement approach is pursued by e.g. Stock and Watson (1991) and Stock and Watson (1993) and is briefly discussed in Section 5.4 on multivariate approaches.

In recent years methods based on likelihood or posterior distributions have been in focus. In the general theory on statistical surveillance there are proofs for their optimality properties (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. The methods are denoted PHM, SRLin and MSR. 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 PHM method is based on a regime switching hidden Markov model (HMM) and 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) and Koskinen and Öller (1998). The SRLin method is derived here by the Shiryaev-Roberts technique (see Section 2.4) under the assumption of known slopes in the respective phases. The MSR method is suggested by Frisé (1994) and evaluated by Andersson (1999) and Andersson (2001). The MSR method is a non-parametric version of the SRLin method. 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, to estimate parameters and the effect of assumptions regarding i) the shape of a turning point and ii) the distribution of the time of transitions.

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 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.

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 (1999). The performance in general surveillance is evaluated using measures such as average run length, expected delay, probability of successful detection and predictive value of an alarm, as suggested by Frisé (1992).

The paper is organized as follows. Section 2 contains a description of different likelihood based approaches, specifically of the PHM, the SRLin and the MSR methods. It also contains theoretical analyses of the effect of some assumptions. In Section 3 the choice of models for the simulation study is motivated and results on the effects of different assumptions are presented. In Section 4 the three methods are used to analyze a period of the Swedish industrial production and the pros and cons of this way to evaluate methods is discussed. In Section 5, the special problems of seasonal data, auto-correlation, trend adjustment and multivariate processes are briefly discussed. Section 6 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, presented at the end of Section 2. The implications of some of these assumptions are discussed in this section, while some are examined by simulation studies in Section 3. Special data problems such as seasonal effects, autocorrelation and multivariate problems are treated in Section 5. The variable considered in Section 2 and 3 is assumed free from seasonal variation, univariate and without autocorrelation (conditional on the regime).

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 for all methods considered here is to detect a change from expansion to recession (or vice versa) in the leading indicator as soon as possible after it has occurred. The inference situation is one of surveillance, i.e. continual observation of a time series with the goal of detecting a turning point in the underlying process as soon as possible. Some of the likelihood based methods use an HMM, to describe the underlying process that changes at an unknown time. Apart from just detecting the change from one phase to another (from expansion to recession), usually an additional aim when using HMM 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. That means that the vocabulary of statistical surveillance is suitable.

2.1 Model within each expansion- and recession phase

Denote the process under observation by X and the observations available at time t by $X_t = \{X(t^*), \dots, X(1), X(2), \dots, X(t)\}$. Time $t=1$ is the first time point in a period of special interest. The model discussed here is:

$$X(t) = \mu(t) + \varepsilon(t), \quad (1)$$

where $\varepsilon(t) \sim \text{iid } N[0; \sigma^2]$
and $\mu(t)$ is a stochastic process to be described below.

$\mu(t)$ can be regarded as a regression function since it is the expectation of $X(t)$ conditional on t . $\mu(t)$ can also be described using an HMM with two states. The aim is to detect a change in μ , from expansion state to recession state. The assumptions in (1) might be too simple for some applications, but are used here to emphasize the inferential issues. These assumptions are the ones, which most suggested methods are based on. Ways to handle models with seasonal effects, autocorrelation and multivariate situations are discussed in Section 5.

By the definition of a turning point, the regression μ is monotonic within each regime. That is

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

where $t=1$ is in a period of expansion

and τ is the turning point from the expansion to a recession.

Observe that the dependency of $E(X(t) | t, \tau) = \mu(t)$ on τ makes $\mu(t)$ a stochastic variable.

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

The PHM method, and many other HMM approaches, relies on the additional assumption that the regression is linear within each phase with parameters which are considered as known. That is

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

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

The SRLin method is also based on the assumption of linearity within phases. In the derivations in Section 2.4.2 it is in addition assumed that the slopes are symmetrical for the two phases. 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 MSR method is studied in Andersson (2001). One aspect of an asymmetry of the business cycle is possible differences in slopes between phases of recession and expansion. The effect of a mis-specification of the slope on the performance of the SRLin method is investigated in Section 3.3.3.

Estimates involved in the alarm statistics are described in Section 2.3 and 2.4. For the MSR method the alarm statistic can be constructed directly from the data to be analyzed. For the PHM method, the slopes are first estimated by data from an earlier period. These estimates are considered as known constants in the derivation of the method for surveillance. For the SRLin method the alarm statistic is constructed under the assumption of known slopes.

The standard deviation σ can be assumed to be different for recession and expansion as in e.g. Koskinen and Öller (1998). Hussey (1992) demonstrates difference for one indicator but not for another. Furthermore, macroeconomic time series are sometimes considered to have a continuously varying standard deviation. 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. The observation X is the logarithm of the original observation. After this transformation, the variance is here assumed equal, as also in Andersson (2001). The surveillance is conducted and evaluated for the transformed variable X .

2.2 Event to be detected

Some approaches, like that of Birchenhall et al. (1999), discuss both prediction and detection of a regime change in a leading index. However, most approaches deal solely with the problem of detecting a change in a leading index, which is considered to predict a change in the business cycle. Only turning point detection in the leading

index will be discussed here. For this problem the detailed specifications differ for different approaches. This will now be described.

For the MSR 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, or vice versa. 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 simplicity, the turning point will be expressed as a peak in the following. That is, the aim is peak detection, i.e. detection of transition from expansion to recession. Thus, for the MSR method the aim is to discriminate between the following two events:

$$\begin{aligned} D_{MSR}(s): \mu(1) \leq \dots \leq \mu(s) \\ C_{MSR}(s): \mu(1) \leq \dots \leq \mu(\tau-1) \text{ and } \mu(\tau-1) \geq \mu(\tau) \geq \dots \geq \mu(s), \\ \text{where } \tau = \{1, 2, \dots, s\} \\ \text{and at least one inequality is strict in the second part.} \end{aligned} \quad (3)$$

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

$$D_{SRLin}(s): \mu(s) = \beta_0 + \beta_1 \cdot s \quad (4)$$

$$\begin{aligned} C_{SRLin}(s) = \{\cup C(\tau)\}, \\ \text{where } C(\tau): \mu(s) = \beta_0 + \beta_1 \cdot (\tau-1) - \beta_1 \cdot (s-\tau+1), \\ \text{and } \tau = \{1, 2, \dots, s\} \text{ and } \beta_0 \text{ and } \beta_1 \text{ are known constants.} \end{aligned}$$

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

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

The difference between the events for SRLin and MSR is only the assumptions on $\mu(t)$. However, for the PHM method the events are different also in another aspect. The apparently simpler event in the PHM approach is combined with a more complicated situation for the information of previous states. No knowledge of previous states is utilized in the PHM expression for the posterior probability. Thus, the probabilities for the history on those earlier states will have an effect. Both the events $D_{PHM}(s)$ and $C_{PHM}(s)$ correspond to families of histories of states. Because of Markov dependence, the earlier observations carry information of the history of states. Thus, $C_{MSR}(s)$ and $C_{SRLin}(s)$ only concern the last turning point, whereas $C_{PHM}(s)$ 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

2.3.1 Assumptions

The probability of a transition from recession to expansion (or vice versa) are, in most approaches in the HMM framework, assumed to be constant with respect to time, see e.g. Layton (1996). But the transition probability can also be assumed time varying, as by Neftci (1982), Diebold et al. (1994) and Filardo (1994). The assumption of a time invariant transition probability is made for all three methods investigated in this paper.

For the PHM method, the information about the time and type (peak or trough) of the last turning point is not utilized. Thus, besides the inference about whether the turning point has occurred, a probability statement has to be made regarding the type of the next turning point. For this purpose, it is necessary to consider all previous possible turns. Therefore, two intensities are needed.

The intensity parameters can be given as Markov transition probabilities (p_{12} and p_{21}) from state 1 to state 2 and vice versa, where state 1 denotes expansion and state 2 denotes recession. This approach is used by Koskinen and Öller (1998). Contrary to the assumption of the PHM method, both the MSR and SRLin method assume that the type of the next turn (peak or trough) is known. This is also assumed by e.g. Neftci (1982). When information about the type of the next turning point is used, it is sufficient with one measure of intensity, v_{12} , if surveillance is made in order to detect a peak or v_{21} if surveillance is made in order to detect a trough. This single transition probability can be written as

$$v = P(C(t)|D(t-1)) = P(\tau = t | \tau \geq t). \quad (6)$$

The duration of a cycle can vary greatly. In order to achieve an estimated transition probability with small enough standard deviation, a very long time series is required. The assumption of a constant transition probability and thus a geometric distribution for the turning point time, τ , is not very realistic in the business cycle application. The lengths of the cycles vary more than for a geometrical distribution and the probability of small values of the time for the turning point is much smaller for the business cycle than for a geometrically distributed variable. Kim and Nelson (1998) investigate duration dependence, that is if the tendency to switch state depends on the time spent in the current state. A test is carried out on four economic indicators in the US economy and the result is that there is a tendency to duration dependence.

The SRLin and MSR methods use the approach of a non-informative prior for the turning point time. Since the likelihood method is optimal for the intensity that is used, the approach with a non-informative prior means that SRLin and MSR are not optimal. However, the non-informative prior also means that these methods are not sensitive to assumptions regarding the intensity.

The transition probabilities might be time dependent. If reliable information about this was known, then very accurate decisions could be achieved. However, the risk is that errors in assumptions would override the information from the data. All approaches described in this paper use transition probabilities which do not depend on time.

The transition probability from one phase to another is estimated from an earlier period and treated as known constants by the PHM method. The transition probabilities will have an effect on the weight that different observations will have in the test statistic. This can be assumed to have a minor influence as long as the estimates are fairly close to the true values. Greater influence can be expected on the alarm rate. By (8) 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 PHM method is sensitive to the values of the transition probabilities.

A non-informative prior is used by the MSR approach and therefore the transition probability does not need to be estimated. When the distribution of τ is unknown, this approach is not optimal, but the approximation works well, even for intensities as high as 0.20 (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). The

Shiryayev-Roberts approach is robust but not exactly optimal. The robustness reduces the risk of systematically wrong inference. Since SRLin and MSR make a minimum of assumptions about the distribution of τ , the methods are not optimal in this respect. However, this also means that the risk of errors due to erroneous assumption or uncertain estimates or assumptions is avoided.

Below the observed sample density function from the estimation period (1970Q1:1987Q1) is compared to geometric density functions using different intensity estimates (see Section 2.3.2).

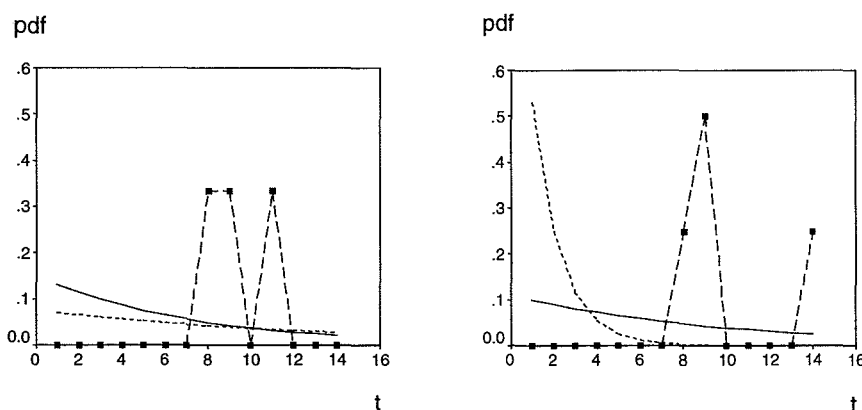


Figure 1: *Left: density for peak with observed sample density function (---■) compared to geometric density functions with intensities 0.07 (···) and 0.13 (—).* *Right: density for trough with observed data (---■) compared to geometric density functions with intensities 0.53(···) and 0.10 (—).*

The geometric distribution is often assumed for τ . When comparing the observed sample density to different geometric distributions above, one can see that the observed density is far from geometric. However if a prior based on observed data, for example with a high weight for a turning point after 10 quarters, was used in the PHM method, the influence of data would be reduced and the probability of an alarm after 10 quarters would be very high.

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 regime dynamics are not modeled. The model is not based on any known or estimated transition probability between the states. Instead, the likelihood approach is used to derive a logistic classification model in which the parameters for different leading indices are estimated. The classification into different regimes is based on explaining variables but not on the earlier state. No dependence between successive events is thus included in their model. One implication of that is that the information on earlier regime types, which by their approach is assumed known, is not fully utilized.

2.3.2 Estimation

Simultaneous maximum likelihood estimation of all parameters in the model is an obvious choice. However, it is pointed out by Lahiri and Wang (1994) and Koskinen and Öller (1998) that if the whole parameter set is estimated using maximum

likelihood the result can be that the rareness of turning points makes large errors around turning points compensated by high accuracy within phases. For that reason, the transition probabilities are sometimes estimated using some other criterion than maximum overall likelihood.

Maximum likelihood estimates based only on the events of transitions are a natural choice. For the data on the Swedish Industrial production, the transition probabilities, p_{12} and p_{21} , are estimated using

$$\hat{p}_{12} = \frac{n_{12}}{n_{11} + n_{12}} = \frac{5}{34 + 5} = 0.13 \quad (0.054), \quad (7)$$

$$\hat{p}_{21} = \frac{n_{21}}{n_{21} + n_{22}} = \frac{4}{4 + 36} = 0.10 \quad (0.047),$$

where n_{ij} is the number of transitions from regime i to regime j and the standard errors are given in parenthesis.

When the posterior probability for PHM is calculated in the simulation study the maximum likelihood estimates $p_{12} = 0.13$ and $p_{21} = 0.10$ of (7) are used.

Some authors, e.g. Koskinen and Öller (1998), use smoothing (see Section 3.3.4) to reduce the stochastic error. Both the transition probabilities and the smoothing parameter are estimated simultaneously from historical information. Koskinen and Öller (1998) suggest that this is done by minimizing a cost-function based on the sum of two measures of error, namely the Brier probability score and an error count estimate. The Brier probability score is the mean square error for the posterior probability, i.e. the average squared deviation between the true state (0 or 1) and the posterior probability. The error count estimate is the proportion of wrongly classified states. From data on the Swedish Industrial production, the resulting transition probabilities are $p_{11}=0.93$ and $p_{22}=0.47$. The value $p_{21}=0.53$ implies that the expected length of a recession, i.e. the expected length before a transition from recession to expansion, is two quarters, which seems unreasonably short. The estimated values $p_{12}=0.07$ and $p_{21}=0.53$ can be compared to the maximum likelihood estimates, based on the same set of data, which are $p_{12}=0.13$ and $p_{21}=0.10$. The result of the ML estimation of p_{21} is very different from the result of the cost function estimation of the same probability. One explanation is that the cost function is used to simultaneously estimate the smoothing parameter (λ , discussed in Section 3.3.4) and the transition parameters p_{12} and p_{21} . The smoothing is made for the purpose of reducing the false alarms. However, a low false alarm rate leads to a low alarm rate generally and a consequence is that also the motivated alarms are delayed by the smoothing. Therefore, if the smoothing parameter is very small (much smoothing), then the transition probability must be very high, in order to reduce the delay. Another explanation to the difficulty to interpret these estimates of the transition probabilities is that the Brier probability score does not take the order of the observations into account. However, the order is crucial for transition probabilities. Thus, the use of the Brier probability score makes the interpretation of the resulting estimates hard.

2.4 Alarm statistics

In all methods discussed here the alarm statistic is based on the likelihood ratio. The 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. The expected delay of an alarm is

minimized for a fixed probability of false alarm. Also other properties of the likelihood ratio method are evaluated and compared with other methods such as the Shewhart method and the CUSUM method, for the case of a shift from an in-control level to an out-of-control level, by Frisé and Wessman (1999).

The LR method gives an alarm for the first time s for which

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

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.

$$\{x_s : P(C|x_s) \geq k\} = \left\{ x_s : \frac{f(x_s|C)}{f(x_s|D)} \geq \frac{P(D) \cdot k}{P(C) \cdot (1-k)} \right\}, \quad (8)$$

where k is the alarm limit for the posterior probability.

The choice of k in (8) is discussed in Section 2.5 on control of false alarms.

2.4.1 PHM

For the PHM method, the alarm statistic at time s is the posterior probability, $P(C|X_s)$. The posterior probability is suggested as alarm statistic by many authors, e.g. Neftci (1982), Hamilton (1989) and Kim and Nelson (1998).

Koskinen and Öller (1998) give the computational formula for the alarm statistic 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

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

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 PHM method, is suitable. However, if only the first change during the evaluation 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 MSR methods.

2.4.2 SRLin

First we derive the LR method which has several optimality properties according to the results of Frisé and de Maré (1991). We derive it for a slightly more general situation than for $C_{SRLin}(s)$ and $D_{SRLin}(s)$ in (4). First we do not require the slopes to be symmetrical, but the events to be discriminated are:

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

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

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

where β_0 , β_1 and δ_1 are known constants.

The optimal alarm rule LRlin for discriminating between $D(s)$ and $C(s)$ above gives an alarm for the first time s where

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

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

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

with $w_j =$

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

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

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

(Details are given in Appendix 1.)

The LRlin(s) statistic is a function of the transition probability $\nu = P(\tau = t | \tau \geq t)$.

The Shiryaev-Roberts 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 method for discriminating between $C_{\text{SRlin}}(s)$ and $D_{\text{SRlin}}(s)$ is given below. The Shiryaev-Roberts approach implies equal weights for the partial likelihood ratios and a constant alarm limit. The alarm rule SRlin for the case of a symmetric turning point with linear functions, using the Shiryaev Robert approach, gives an alarm for the first time s where

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

$$\text{where } w_j = (4 \cdot \beta_1^2 \cdot (j-1) + 4 \cdot \beta_0 \cdot \beta_1) \cdot \sum_{u=j}^s (u-j+1),$$

and L is a constant alarm limit.

The SRlin(s) can be compared to the corresponding Shiryaev-Roberts method (SRcon) for detecting a change from an in-control level (μ^0) to an out-of-control level (μ^1):

$$\sum_{j=1}^s \exp \left[\frac{(\mu^1 - \mu^0)}{\sigma^2} \sum_{u=j}^s x(u) \right] \cdot w_j(s, \mu^1, \mu^0) > k',$$

$$\text{where } w_j(s, \mu^1, \mu^0) = \exp \left[\left(\frac{1}{2\sigma^2} \right) \left((s-j+1) \cdot ((\mu^0)^2 - (\mu^1)^2) \right) \right]$$

and k' is a constant alarm limit.

One major difference between SRLin and SRcon is that the weights for late observations are larger for the SRLin method, which is expected as these will have a greater difference to the D -alternative than those just after the change.

2.4.3 MSR

The maximum likelihood ratio statistic at time s is

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

where $C = \{\tau \leq s\} = \{\tau=1, \dots, \tau=s\}$ and $D = \{\tau > s\}$.

The event $D = D_{MSR}$ implies a monotonically increasing μ -vector $\{\mu(t) \leq \mu(t+1)\}, t \geq 1$.

The denominator of $\text{MLR}(s)$ 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 $C = C_{MSR}$ we have $C = \{C_1, C_2, \dots, C_s\}$, where C_j implies $\{\mu(1) \leq \dots \leq \mu(j-1), \mu(j-1) \geq \mu(j) \geq \dots\}, j \in \{1, 2, \dots, s\}$.

In the numerator of $\text{MLR}(s)$ 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 is 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 MSR method gives an alarm for the first s for which

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

exceeds a fixed limit. The method is suggested in Frisén (1994) and evaluated by Andersson (1999) and Andersson (2001).

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 (1999) suggest that the control is made by a statistic similar to the ARL^0 , namely the MRL^0 , which is the median run length. This has several advantages, such as easier interpretations for the skewed distributions and much shorter computer time for calculations. Here, the median, $M[t_A]$, is defined to be $m_1 + (0.5 - p_1)/(p_2 - p_1)$, where $P(t_A \leq m_1) = p_1$, $p_1 < 0.50$ and $P(t_A \leq m_1 + 1) = p_2$, $p_2 > 0.5$.

The time of the alarm, t_A , for the SRLin and MSR methods are

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

and

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

respectively, where k_{MSR} and k_{SRLin} are constant alarm limits, determined to yield $M[t_A | \tau = \infty] = MRL^0$ for both methods, where MRL^0 is a chosen constant.

A more direct Bayesian approach, which is often used, is to control the limit for the posterior probability. This approach is also used for PHM. The time of the alarm, t_A , for the PHM method is defined as

$$t_A = \min [s: P(C|X_s) > 0.5]. \quad (11)$$

The limit 0.5 for the posterior probability is used also by other authors, e.g. Hamilton (1989) and Koskinen and Öller (1998). 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

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

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, Frisé (1999) demonstrates that the price is high. The ability to detect a change will be very low if it happens a long time after the monitoring has started.

2.6 Assumptions on knowledge of the type of the next turn

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 MSR approach, nothing will be gained by including earlier time points in the analyses. However, without this information, 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 PHM method and by e.g. Koskinen and Öller (1998). The latter approach is used for the SRlin and MSR methods and by e.g. Neftci (1982) and Diebold and Rudebusch (1989). By comparisons between 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 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. PHM.) 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. An illustration of this by a demonstration of how the knowledge about the type of the next turning point is implemented in the HMM approach for decision time $s=3$ is given in Appendix 2, where also the effect on the posterior probabilities for recession or expansion is demonstrated.

2.7 Assumptions on knowledge of the regime of all earlier observations

Past periods with known regime characteristics carry valuable information. Several authors utilize this information for estimation purposes.

Birchenhall et al. (1999) assume knowledge about earlier regimes. However, as was seen above, one implication of their approach is that the information on earlier regime types is not used. In their work, data from earlier time points are used for selection and estimation of the weights for the different leading indicators of the classification rule. Koskinen and Öller (1998) model the dynamics of the transitions but still use the knowledge of the regime of prior observations only for estimation purposes. We will now demonstrate that a Shewhart type of method (only the last observation is used) will be optimal for the PHM setting in this situation.

The alarm rule for the PHM method is

$$\frac{P(C(s)|y(s-1)) \cdot f(y(s)|C(s))}{f(y(s)|y(s-1))} > k, \text{ where } k = 0.5.$$

This is equivalent to

$$\frac{P(C(s)|y(s-1)) \cdot f(y(s)|C(s))}{P(C(s)|y(s-1)) \cdot f(y(s)|C(s)) + P(D(s)|y(s-1)) \cdot f(y(s)|D(s))} > k,$$

where the following calculations are used:

$$P(C(s)|y(s-1)) = (p_{12} \cdot P(D(s-1)|y(s-1)) + p_{22} \cdot P(C(s-1)|y(s-1)))$$

and

$$P(D(s)|y(s-1)) = (p_{11} \cdot P(D(s-1)|y(s-1)) + p_{21} \cdot P(C(s-1)|y(s-1))),$$

where p_{12} , p_{22} , p_{11} , p_{21} are transition probabilities.

If the regime of the previous observation, $y(s-1)$, was known and equal to C , i.e. if

$$P(C(s-1)|y(s-1)) = 1$$

then

$$p_{21} = P(D(s)|C(s-1)) = P(D(s)) \text{ and } p_{22} = P(C(s)|C(s-1)) = P(C(s))$$

and therefore the alarm rule for the PHM method is reduced to

$$P(C(s)|y(s)) > k.$$

Analogously, it can be shown that if the regime of the previous observation, $y(s-1)$, was known and equal to D , then the alarm rule for the PHM method is again reduced to

$$P(C(s)|y(s)) > k.$$

Thus, for the situation where the regimes for all past time points are known, the optimal alarm statistic is based only on the last observation, $y(s)$.

Table 1. *Summary of the assumptions used for some methods.*

	Neftci (1982)	Hamilton (1989)	Birchenhall (1999)	PHM	SRLin	MSR
Type of next turn used	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
Alarm rule	$P(C x) > 1 - P(t_A < \tau)$	$P(C x) > 0.5$	$P(C x) > 0.5$ and $P(C x) > \hat{p}$	$P(C X) > 0.5$	$P(C X) > k_{SRLin};$ MRL^0 fix	$P(C X) > k_{MSR};$ MRL^0 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

3 MONTE CARLO STUDY ON THE EFFECTS OF DIFFERENT SPECIFICATIONS

3.1 Model for the evaluation period

The investigation of the effects of different specifications is made for the detection of a turn in an evaluation period with one turn. In this 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 time series is illustrated in Figure 2, raw data and after adjustment for seasonality. A large part of the IP series is used in the estimation process for the PHM method and a part of the IP series that contains one turning point is used in the evaluation.

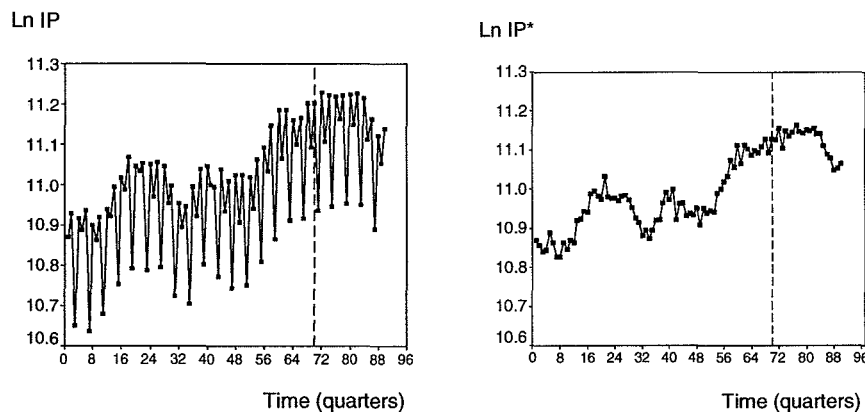


Figure 2: *Industrial production, quarterly data (1970Q1: 1992Q2). The evaluation period starts at 1987Q2, marked with a dashed vertical line. Left: raw data, right: seasonally adjusted data.*

The expansions and recessions are dated using the records of the National Institute of Economic Research (1992). In the evaluation study a model is used that resembles quarterly seasonally adjusted data on IP for the period 1987Q2:1992Q2, which includes one peak and where the date 87Q1 is defined as a trough by the National Institute of Economic Research. The seasonal adjustment is made using regression on seasonal dummies. Whether the evaluation period starts at the last turn in the estimation period, or a few quarters after the last turning point, has a great impact on the false alarm probability of the PHM method, especially for the first time point. This effect is investigated in Section 3.3.2. In a realistic situation, the knowledge that a time point is a turning point cannot be confirmed directly. It is reasonable to think that the confirmation can come after 3 or 4 quarters. Therefore, in the simulation study, unless anything else is specified, the evaluation period ($t=1$) starts 4 time points after the last turning point of the estimation period.

The procedure to determine the models, used in the simulation study, will now be described.

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

$$\begin{aligned} X^D(t) &= \mu^D(t) + \varepsilon(t), \\ \text{where } \mu^D(t) &= 11.194 + 0.0069 \cdot t, \\ \text{and } \varepsilon(t) &\sim \text{iid } N[0; 0.016]. \end{aligned} \tag{12}$$

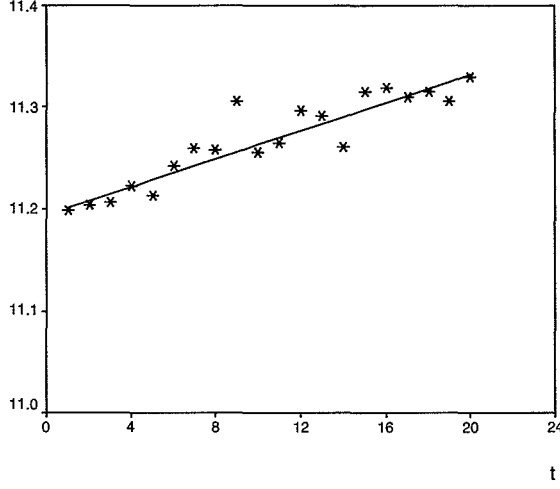


Figure 3: The regression line (—) for event D (no turn) and one realization (*).

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), which starts at a trough. A regression curve, which is piecewise of the third degree and piecewise of the first degree, is found to fit well. The seasonal effects are included as seasonal dummies when the parameters of the regression curve and the standard deviation are estimated. Thus, the model for the variable X is

$$\begin{aligned} X^C(t) &= \mu^C(t) + \varepsilon(t), \\ \text{where} \\ \mu^C(t) &= \begin{cases} 11.194 + 0.0066 \cdot t - 0.00017 \cdot t^2 - 0.000015 \cdot t^3, & 1 \leq t \leq 13 \\ 11.340 - 0.0089 \cdot t, & t \geq 14 \end{cases} \\ \varepsilon(t) &\sim \text{iid } N[0; 0.016]. \end{aligned} \tag{13}$$

The regression has a peak at time $t=10$ and thus, 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.

The rounded curve described above 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.

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

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

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

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

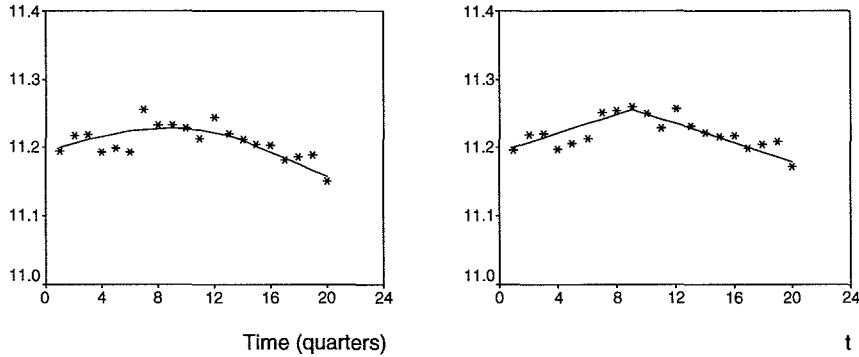


Figure 4: The regression curve (—) for a turn at 10 and one realization (*). Left: the rounded curve. Right: the piecewise linear curve.

3.2 Other specifications needed for the simulation procedure

3.2.1 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 Koskinen and Öller (1998)) and here also by the PHM method. One object in our simulation study is to study the effect of estimation of parameters in the regression function. It is often suggested that also other parameters, e.g. the transition probabilities are estimated. The effects of errors in the estimated values are discussed but no simulation study is made for those parameters.

The PHM method estimates its parameters (μ and σ for recession and expansion, respectively) from an estimation period. In order to incorporate the variation introduced by this use of previous stochastic data, also these values are simulated. The estimation period is 1970Q1 to 1987Q1, see Figure 2. This period includes several peaks and troughs. The dates of these turning points, given by the National Institute of Economic Research (1992), are used in developing a simulation model for the estimation period. 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 intercept of the regression models is 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. Observations on X are thus simulated according to

$$X_i(t) = \mu_i(t) + \varepsilon_i(t), \quad (15)$$

$$\text{where } \mu_i(t) = \begin{cases} \mu_{1j}(t), & j = \{1, 2, 3\} \\ \mu_{2j}(t), & j = \{1, 2, 3, 4\} \end{cases}$$

$$\text{and } \varepsilon_i(t) \sim iid \begin{cases} N(0; \sigma_{1j}^2), & j = \{1, 2, 3\} \\ N(0; \sigma_{2j}^2), & j = \{1, 2, 3, 4\}. \end{cases}$$

The function $\mu_{1j}(t)$ and parameter σ_{1j} represent the expected value and the standard deviation in expansion phase j and $\mu_{2j}(t)$ and σ_{2j} represent the expected value and the standard deviation in recession phase j . The μ -function used in the simulation of the estimation period is given in the Appendix 3 and is illustrated in Figure 5.

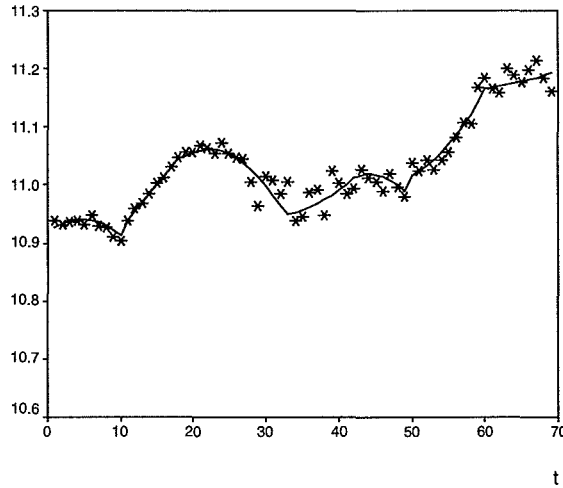


Figure 5: The expected value (—) and one realization (*), for the model used to simulate the estimation period (1970Q1:1987Q1).

3.2.2 Control of false alarms

For the PHM 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 MSR 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 MSR and 0.12 for SRLin. The standard errors are estimated in a simulation study, using the empirical distribution.

3.3. Evaluation of the effect of different specifications

The evaluation and comparison 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, see Frisén (1992).

3.3.1 Comparison between the three specific methods

One main difference concerns the assumption of regression on time. Two of the methods, SRLin and PHM, 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 PHM) and has parameter values $\beta_1 = 0.0069$, $\sigma = 0.016$ (which is assumed by SRLin). A study on the effect of errors in these assumptions is conducted in Section 3.3.3. In the following sections we will examine which of the other differences in specifications that have a major impact.

3.3.1.1 False alarm

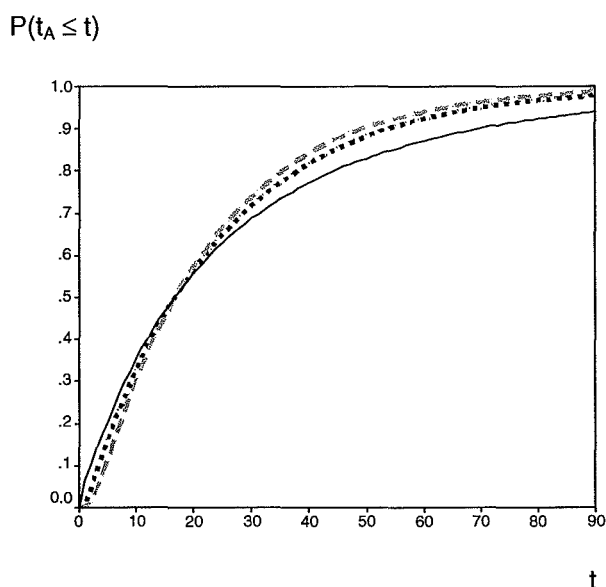


Figure 6: The distribution of the time of an alarm conditional on event D (no turn). PHM (—), MSR (---), SRLin (···).

As is seen in Figure 6, the PHM method has more frequent false alarms at early time points, but low alarm probability later compared with that of the SRLin and especially MSR. The curves cross at run length 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 $MRL^0 = M[t_A | \tau = \infty] = 17$. The curve for the SRLin method is between the other two.

The conditional probability of a false alarm, $P(t_A < \tau | \tau = 10)$, for the rounded curve in Figure 4, is 0.46 (0.0035) for PHM and 0.41 (0.0033) for MSR. For the corresponding piecewise linear case in Figure 4, $P(t_A < \tau | \tau = 10)$ is 0.33 (0.0023) for PHM, 0.27 (0.0022) for MSR and 0.29 (0.0023) for SRLin. The higher false alarm probability for the rounded curve in Figure 4 is due to a decreasing slope just before the turn. The conditional probabilities of a false alarm for different values of τ are summarized by the total probability of a false alarm, $P(t_A < \tau)$, which is the expected value of $P(t_A < \tau | \tau = t)$. $P(t_A < \tau)$, as a function of ν , is presented in Figure 7 for the case when τ has a geometric distribution with intensity ν .

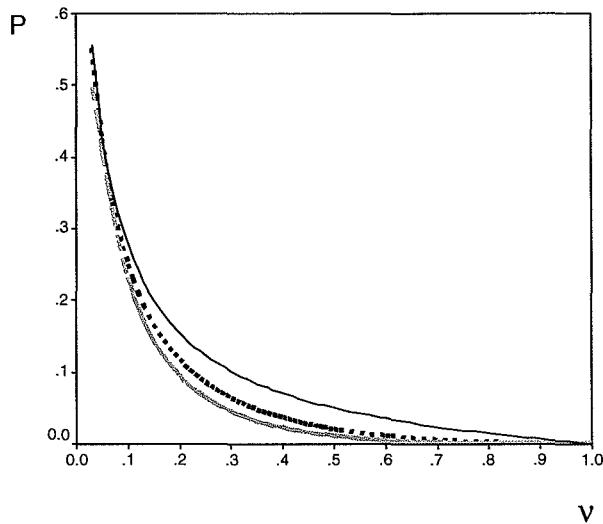


Figure 7: The probability of a false alarm when τ is geometrically distributed with parameter v . PHM (—), MSR (---), SRlin (···).

When comparing PHM, MSR and SRlin we see that the false alarm probability is the smallest for the MSR and the largest for the PHM for every value of the intensity, v . 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 PHM for all v is a result of the error-spending curve of PHM, with many early alarms. The dramatic decrease as v tends to one is because the expected value of τ decreases and most of the alarms are thus triggered after the expected turning point.

3.3.1.2 Delay of a motivated 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 8 for the three methods below.

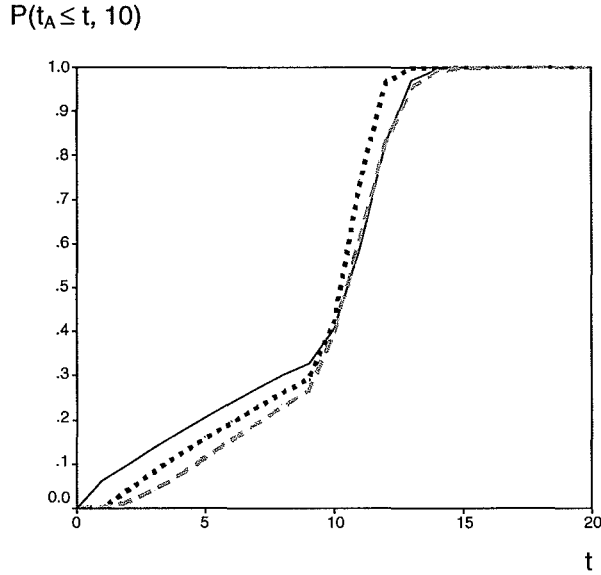


Figure 8: The distribution of the time of an alarm for the piecewise linear curve, $\tau=10$. PHM (—), MSR (---), SRlin (···).

The evaluation of the ability to detect an event C (a peak) is made using the conditional expected delay time of an alarm, $CED(\tau) = E[t_A - \tau | t_A \geq \tau]$, and the conditional median delay, $CMD(\tau) = M[t_A - \tau | t_A \geq \tau]$.

Table 2: Conditional expected delay and conditional median delay for a turn at 10.
 $\tau = 10$

	Rounded		Piecewise linear	
	CED	CMD	CED	CMD
PHM	1.50 (0.012)	0.86	1.79 (0.0065)	1.30
MSR	1.64 (0.011)	1.05	1.65 (0.0068)	1.08
SRlin	--	--	1.23 (0.0048)	0.70

The delay times for the piecewise linear case in Figure 4 are summarized by $CED(\tau)$ and $CMD(\tau)$ in Figure 9 and 10, respectively. For $\tau \leq 20$ we have $sd[\hat{C}ED(\tau)] \leq 0.0079$ for PHM, $sd[\hat{C}ED(\tau)] \leq 0.0087$ for MSR and $sd[\hat{C}ED(\tau)] \leq 0.0059$ for SRlin.

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

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

Using $\nu=0.10$ in the geometric distribution, ED is 1.79 (0.0024) for PHM, 1.99 (0.0026) for MSR and 1.26 (0.0018) for SRlin.

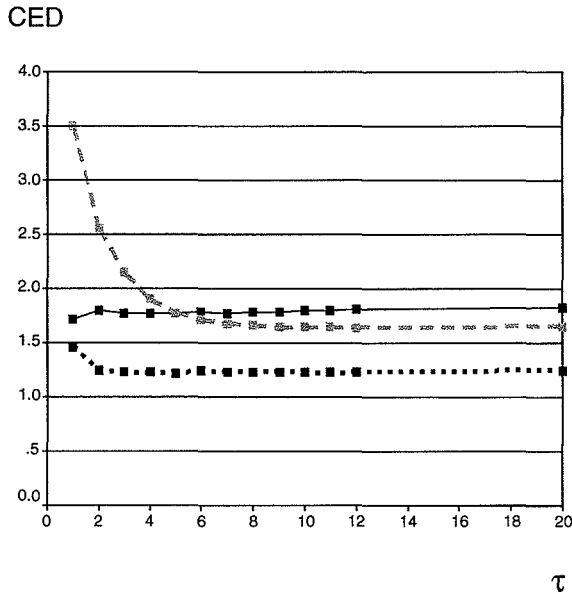


Figure 9: Conditional expected delay for the piecewise linear curve with turn at τ . PHM (—■), MSR (---■), SRLin (···■).

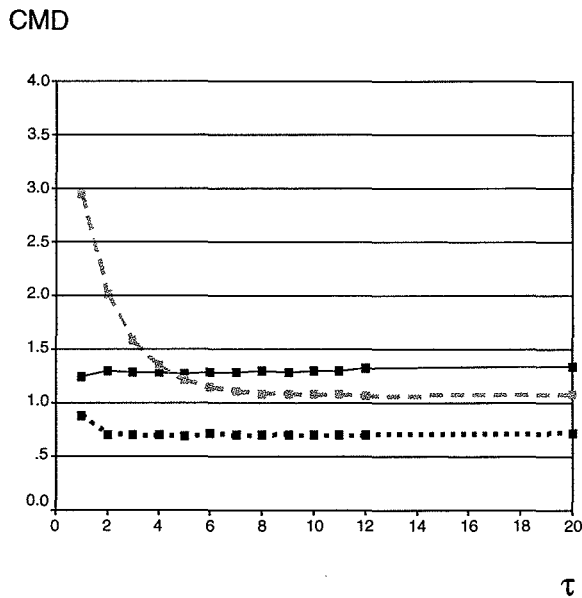


Figure 10: Conditional median delay for the piecewise linear curve with turn at τ . PHM (—■), MSR (---■), SRLin (···■).

When comparing PHM, and MSR we see that both the conditional expected delay and the conditional median delay is worse for MSR, for small values of τ , $\tau < 4$. After that, the delay is slightly shorter for MSR, compared to PHM. The effect of τ is large for MSR for small values of τ . However, an asymptote is reached at about $\tau=10$. For PHM, the effect of τ is very small. A very slight increase in both conditional expected delay and conditional median delay can be observed in Figure 9 and 10, as τ increases. The SRLin method has the shortest delay for every τ . Both SRLin and PHM reach their respective asymptote already at 2. The reason is that both these methods

assume the correct parametric function for the turn (a piecewise linear function). The MSR method needs more observations in the beginning to have enough evidence of a turn.

3.3.1.3 Probability of successful detection

The probability of successful detection within d time points, $\text{PSD} = P((t_A - \tau) \leq d | t_A > \tau = \tau_0)$, is given in Figure 11 for $\tau=10$. For the rounded curve in Figure 4 we have that $\text{sd}[\text{PSD}] \leq 0.0048$ for PHM and $\text{sd}[\text{PSD}] \leq 0.0044$ for MSR. For the piecewise linear curve in Figure 4 we have that $\text{sd}[\text{PSD}] \leq 0.0030$ for PHM, $\text{sd}[\text{PSD}] \leq 0.0029$ for MSR and $\text{sd}[\text{PSD}] \leq 0.0029$ for SRlin.

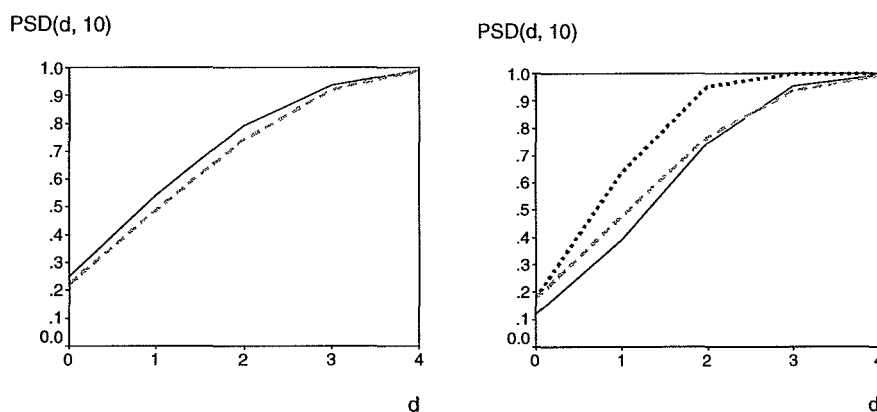


Figure 11: Probability of successful detection within d time points for $\tau=10$. Left: the rounded curve, PHM (—), MSR (---). Right: the piecewise linear curve, PHM (—), MSR (---), SRlin (···).

The PSD curves are very similar for the PHM and MSR methods. For MSR it is showed that the PSD increases as the post peak slope grows steeper (see Andersson (2001)). 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.

3.3.1.4 Predictive value of an alarm

Another evaluation measure is the predictive value of an alarm at time t , $\text{PV}(t) = P(\tau \leq t | t_A = t)$. This reflects the trust you should have in an alarm. In Figure 12 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 used and for $t = \{2, \dots, 12\}$, simulated values are used. From Figure 12 it is evident that the price for the high alarm probability in the first point for the PHM method is that those alarms are of little value. Since the predictive value is only 0.2, an alarm would hardly motivate any action. $\text{PV}(1)$ is very high for MSR, as a result of the very low false alarm rate at $t=1$.

PV

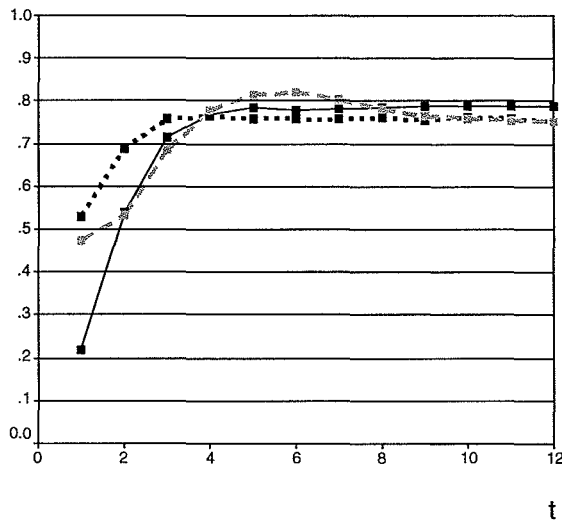


Figure 12: Predictive value of an alarm at time t for $v = 0.1$. PHM (— ■), MSR (- - - ■), SRLin (· · · ■).

Both SRLin and PHM reach their respective asymptote early. The development for MSR is a little different. The predictive value of MSR increases until $t = 6$, after that the predictive value decreases slightly and reaches the same asymptote as SRLin at approximately $t = 10$. The MSR 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 $t = 10$, MSR has the same predictive value as SRLin.

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

3.3.2 Start of the evaluation period

For the methods MSR 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 PHM 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 through 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 large false alarm probability at $t=1$ for the PHM method.

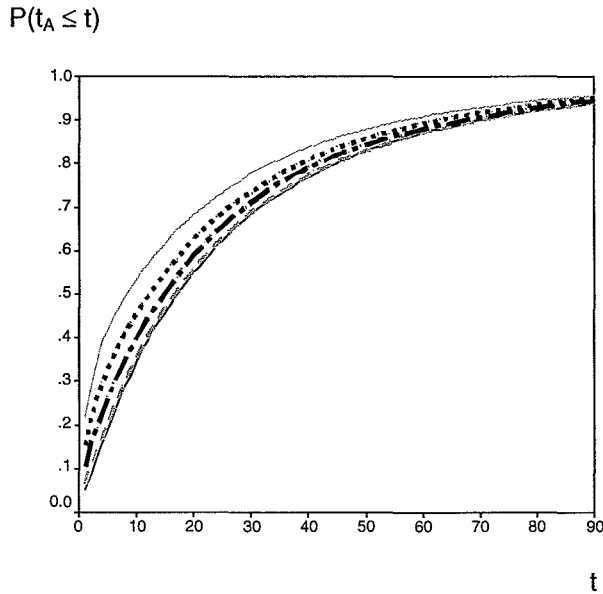


Figure 13: The distribution of the time of an alarm conditional on event D (no turn) for the PHM method. The evaluation period starts $h=\{0,1, 2, 4, 6\}$ time points after the latest turning point, $h=0$ (—), $h=1$ (···), $h=2$ (-·-·-), $h=4$ (- - -), $h=6$ (—).

The false alarm probability at the first time point depends highly on the start of the evaluation period, as illustrated in Table 3 below.

Table 3: False alarm probability, at the first time point, as a function of the start of the estimation period, h . Standard errors are given in parenthesis.

h	$P(t_A = 1 \mid D, h)$
0	0.22 (0.0029)
1	0.15 (0.0025)
2	0.10 (0.0022)
4	0.064 (0.0012)
5	0.057 (0.0016)
6	0.054 (0.0016)

The starting time $h=4$ was used in the simulations if not otherwise stated.

3.3.3 Effect of the estimation procedure for the regression coefficients

If a short period of estimation of the parameters is used, then a considerable variation by the use of these parameters will be introduced. Thus, estimation from short periods will introduce an extra variability and should be avoided. 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. However, if a very long and stable period is used for estimation of the parameters, then the parameters can be considered as known. The estimation procedures described in Appendix 4 will result in a stochastic deviation from the relevant values. Neither SRLin nor PHM incorporates this uncertainty in the methods. If a large set of previous data is available and if it can be assumed that the period to be analyzed has the same pattern as the previous periods, then the regression coefficient can be considered known. The effect

of 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 models for the observations are presented in (12) and (14).

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

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

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

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

We will give results for the specific situation where $\beta_1=0.0069$, $\text{sd}[\hat{\beta}_1]=0.0009$ and $\gamma = \{0.0018, -0.0018\}$. These examples of mis-specification are chosen since they represent values between which approximately 95% of the expansion estimates would be, with the estimation procedure described in Appendix 4. Observe that the mis-specification is small, for early times, compared with the stochastic variation $\sigma=0.016$ around the curve. The correct case and the two mis-specifications are illustrated in Figure 14.

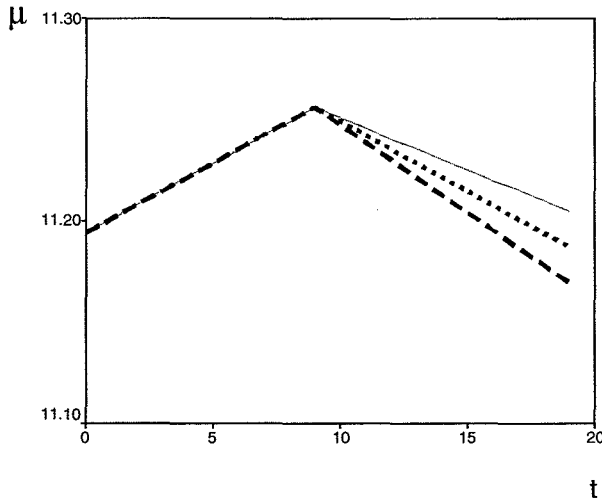


Figure 14: The piecewise linear regression curve for a turn at 10. The post-peak slope is correctly specified (\cdots), too flat (---) and too steep (---).

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

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

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 great. However, when the alarm limit is adjusted to give the same MRL⁰, the situation changes. For this case, (but not the next that will be studied) the adjustment is minor.

The exact alarm probabilities at time 1, using the alarm limit from the simulations, are given in Table 4, for the cases $\{\tau = \infty\}$ and $\{\tau = 1\}$. The value of β_0 corresponds to a translation and has no effect on the probability.

Table 4: Alarm probability at $t = 1$, conditional on no turn ($\tau = \infty$) and turn at 1 ($\tau = 1$).

		P(alarm at time 1)		
		$\beta_1 = 0.0069, \sigma = 0.016$		
		$\gamma = 0$	$\gamma = 0.0018$	$\gamma = -0.0018$
$\tau = \infty$		0.0026	0.0062	0.0009
$\tau = 1$		0.0268	0.0495	0.0107

The densities for the alarm time, when there is still expansion (event D) is shown in Figure 15 below.

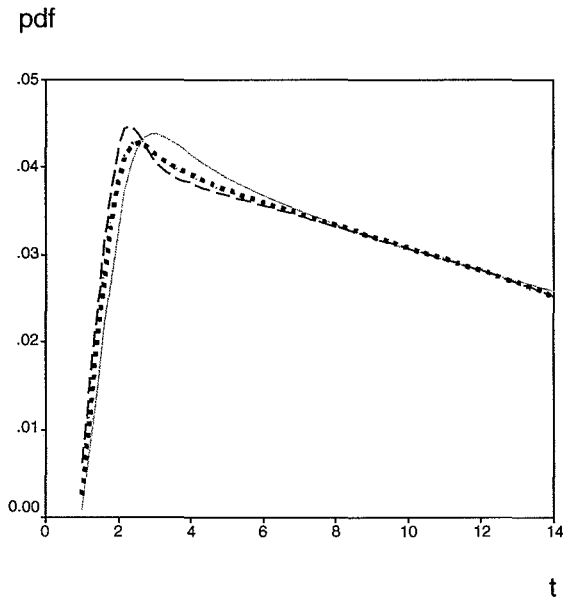


Figure 15: *The density of the time of an alarm conditional on event D (no turn). The post-peak slope is correctly specified (···), too flat (—) and too steep (---).*

We see that the alarms have a tendency to come later when the methods are optimized to detect a small change with a fixed MRL^0 . This is in agreement with the results by Frisé and Wessman (1999) for a shift in level and a fixed ARL^0 . For the values $\gamma \in \{0, 0.0018, -0.0018\}$ no great differences are seen except at the first time points.

The likelihood ratio is optimal for the value of the slope that is used, that is when using $(\beta_1 + \gamma)$ the LR method is optimal for detecting a recession with slope $(\beta_1 + \gamma)$. You would expect that it is wise to save the alarm power until you have gathered much information if you have a hard case (small change) to detect. The result agrees with this. In Frisé and Wessman (1999) the effect of wrong specification in the situation of a change from an in-control level to an out-of control level is investigated. They prove that when the change, for which the method is optimized, tends to infinity then all examined methods tend to the properties of the Shewhart method, which has a geometrical false alarm density.

The expected delay for different values of τ is given below.

CED

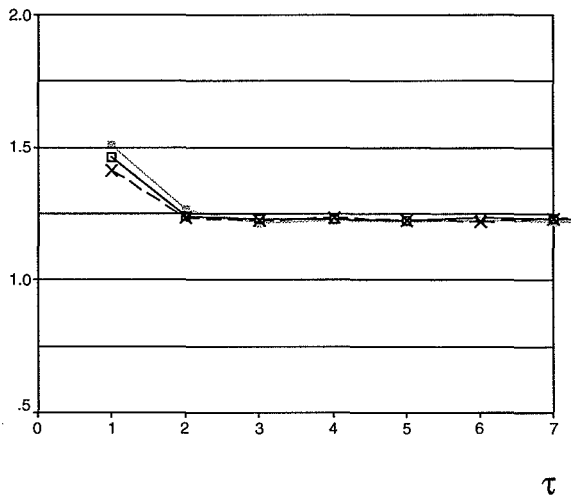


Figure 16: Conditional expected delay for a turn at τ . The post-peak slope is correctly specified (— \square), too flat (— \blacksquare) and too steep (— \times).

No dramatic difference can be detected 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 incorrectly specified slope is small at the beginning of the recession. Since the delays are expected to be small, the mis-specification in slope has a minor effect. For that reason we would not expect any large differences between SRLin with a correctly specified slope and SRLin with an incorrectly specified slope. We have that $sd[\hat{CED}(\tau)] \leq 0.0059$ for $\gamma=0$, $sd[\hat{CED}(\tau)] \leq 0.0056$ for $\gamma=0.0018$ and $sd[\hat{CED}(\tau)] \leq 0.0054$ for $\gamma=-0.0018$.

PV

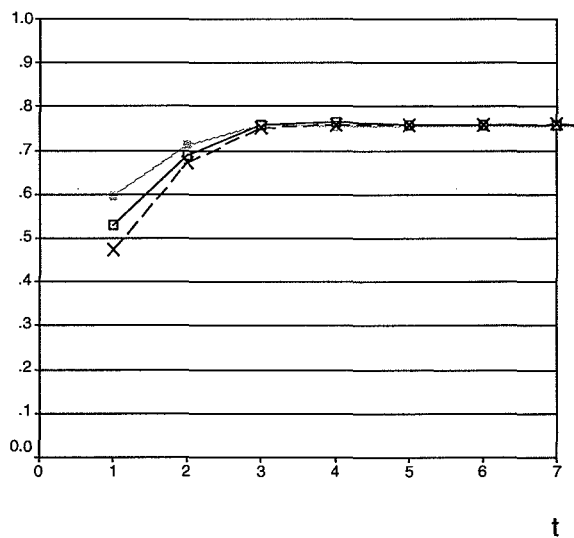


Figure 17: Predictive value of an alarm at time t for $\nu = 0.1$. The post-peak slope is correctly specified (— \square), too flat (— \blacksquare) and too steep (— \times).

The predictive value is very similar between $\gamma=0$ and $\gamma=\{0.0018, -0.0018\}$, 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 is not self-evident for an investigation of the effect of incorrect specification in both D and C . However the case will suffice to demonstrate the dramatic difference compared with the earlier case with incorrect specification only in C .

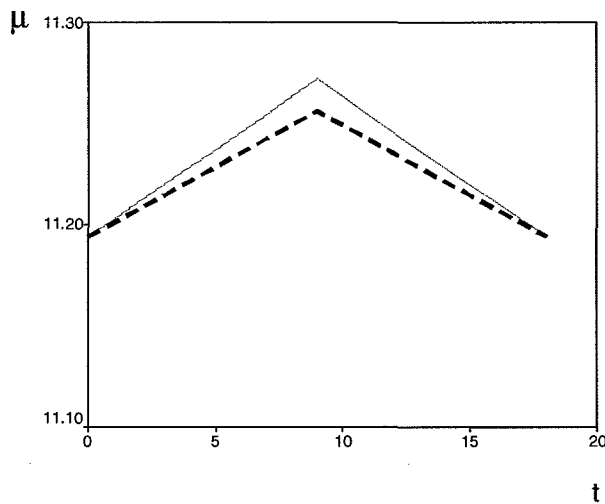


Figure 18: The piecewise linear regression curve for a turn at $t = 10$. Both pre-peak and post-peak slopes are correctly specified (---), too steep (—).

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 studied 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.

The low false alarm probability for small t for the incorrect specification ($\gamma = 0.0018$) in Figure 19 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.

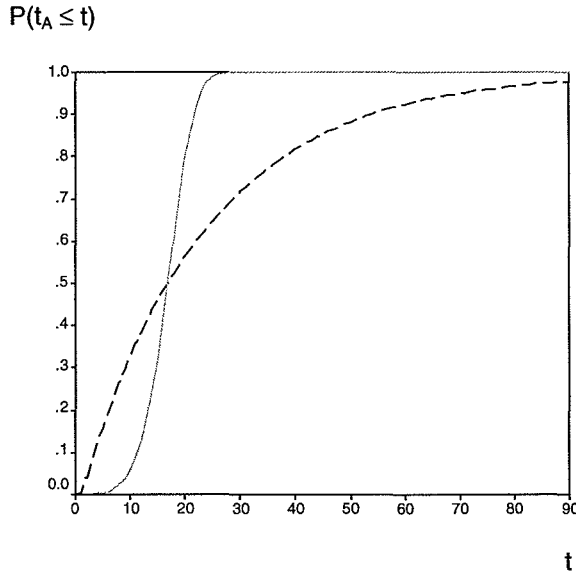


Figure 19: *The distribution of the time of an alarm, conditional on event D (no turn). Both pre-peak and post-peak slopes are correctly specified (---), too steep (—).*

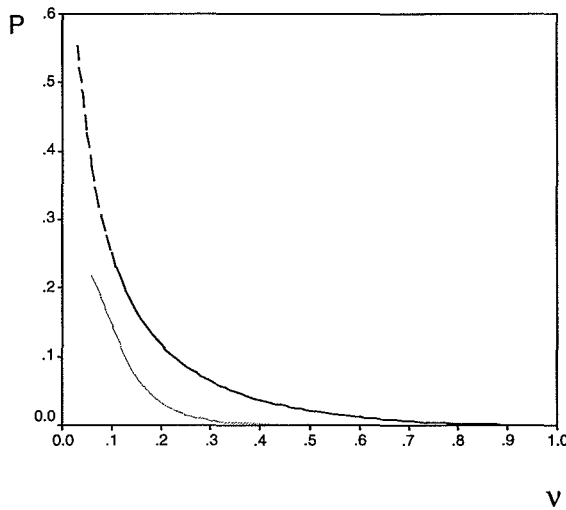


Figure 20: *The probability of a false alarm when τ is geometrically distributed with the parameter v . Both pre-peak and post-peak slopes are correctly specified (---), too steep (—).*

The false alarm probability is the smallest when using the incorrect specified slopes for every value of the intensity, v . 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.

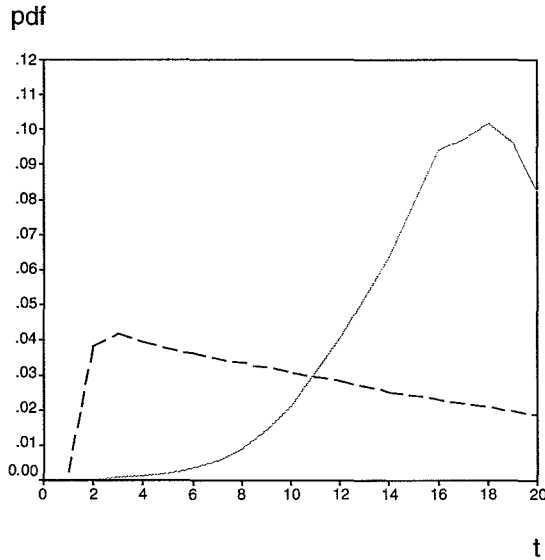


Figure 21: The density of the time of an alarm, conditional on event D (no turn). Both pre-peak and post-peak slopes are correctly specified (---), too steep (—).

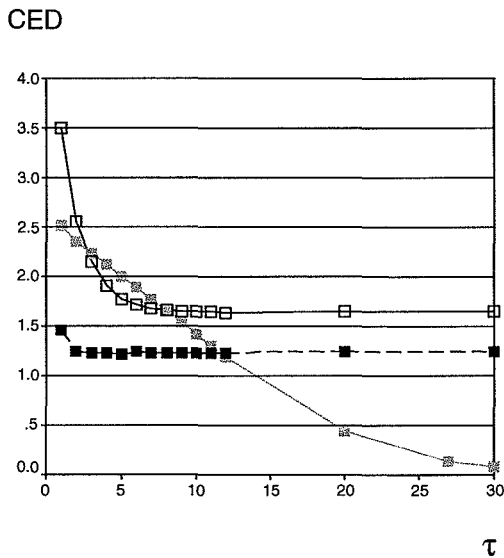


Figure 22: Conditional expected delay for a turn at τ . SRlin correctly specified slopes (--- ■), SRlin too steep slopes (— ■), MSR(— □).

For small values of τ , the conditional expected delay is longer if both slopes (expansion and recession) are over-estimated, as seen in Figure 22. 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, MSR, has a long delay time for early turns, it quickly reaches a reasonable asymptotic value of CED. We have $\text{sd}[\hat{C}ED(\tau)] \leq 0.0059$ for SRlin (correct), $\text{sd}[\hat{C}ED(\tau)] \leq 0.0046$ for SRlin (wrong) and $\text{sd}[\hat{C}ED(\tau)] \leq 0.0087$ for MSR.

The predictive value of an alarm, under the assumption of a geometric distribution with intensity ν is shown in Figure 23.

PV

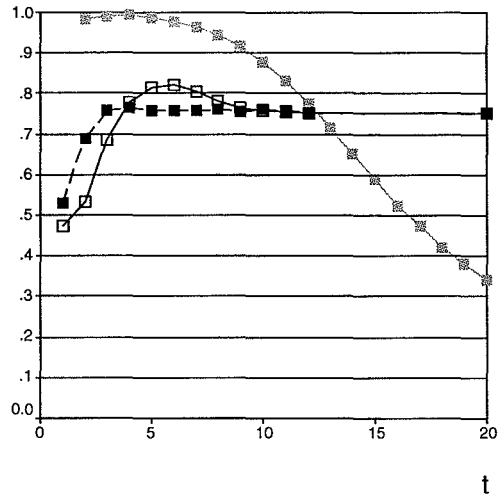


Figure 23: Predictive value of an alarm at time t for $\nu = 0.1$. SRlin correctly specified slopes (--- ■), SRlin too steep slopes (—■), MSR(—□).

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 MSR 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.

3.3.4 Effect of smoothing

In the paper by Koskinen and Öller (1998) the recommendation is that the observations should be smoothed after differentiation, see also Öller (1986). Koskinen and Öller (1998) state that the objective of the smoothing is both reduction of white noise and lagging of the turning point. The latter purpose is motivated by the use of multivariate data, where the turning points of the different processes are not always synchronized. The differentiated observations $y(t)$ are smoothed according to

$$\tilde{Y}(t) = \lambda Y(t) + (1 - \lambda)\tilde{Y}(t - 1),$$

where $\lambda \in \{0, 1\}$.

Smoothing by kernel estimators is used by e.g. Hall et al. (1995). The smoothing of observations reduces the variance and hence reduces the false alarm probability. However, there are also disadvantages. The smoothing will introduce an auto-correlation and reduce the distinctness of the turning point. Alternatives to the smoothing are discussed in Section 5.2. The effect of smoothing on the PHM method is shown below for adjusted alarm limits to give the same MRL^0 .

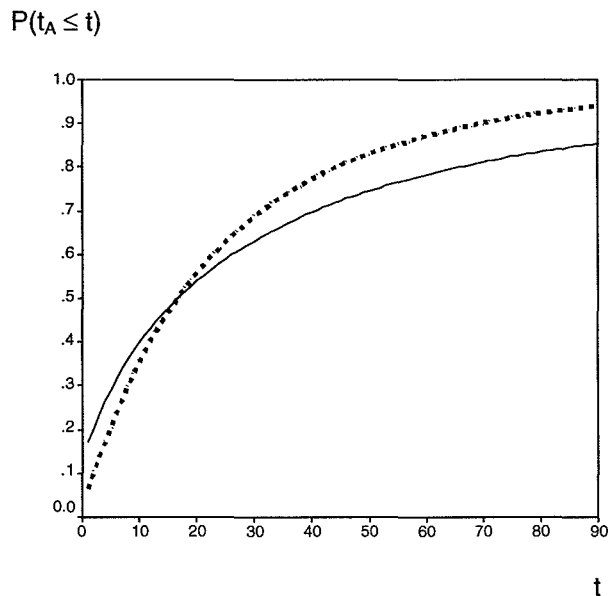


Figure 24: The distribution of the time of an alarm, conditional on event D (no turn). PHM method, $\lambda=1$ (\cdots) and $\lambda=0.3$ (---).

The common value of the MRL^0 is 17. For the case when $\lambda=1$ we have that $sd[MRL^0] = 0.13$ and when $\lambda=0.3$ we have that $sd[MRL^0] = 0.15$. The differences seen in Figure 24 are due to different skewness of the density of the time of the alarm. This difference in skewness is due to differences in autocorrelation and variance.

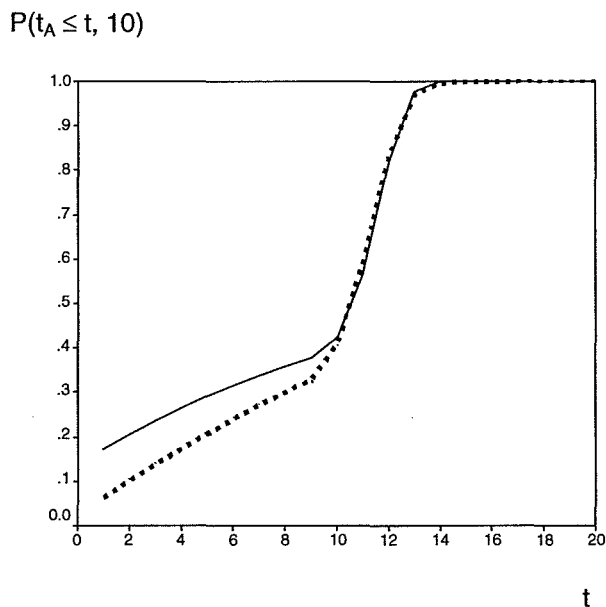


Figure 25: The distribution of the time of an alarm for the piecewise linear curve, $\tau=10$. PHM method, $\lambda=1$ (\cdots) and $\lambda=0.3$ (---).

PSD(d, 10)

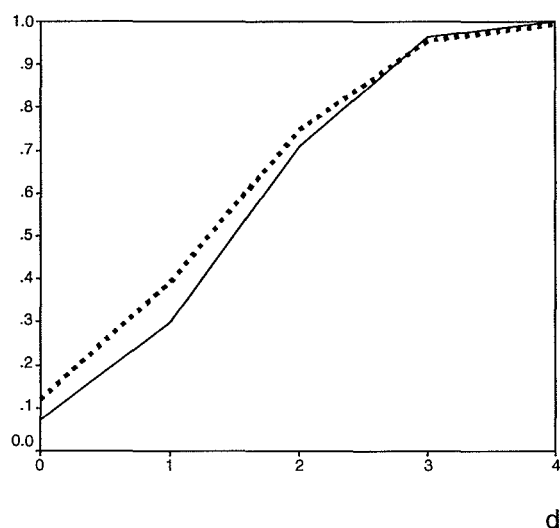


Figure 26: Probability of successful detection within d time points for the piecewise linear curve, $\tau=10$. PHM method, $\lambda=1$ (\cdots) and $\lambda=0.3$ (—).

For the case when $\lambda=1$ we have that $\text{sd}[\text{PSD}] \leq 0.0030$ and when $\lambda=0.3$ we have that $\text{sd}[\text{PSD}] \leq 0.0029$. The reduced distinctness of the turning point, due to smoothing, decreases the probability of successful detection for $\tau=10$.

4 EVALUATION BY ACTUAL DATA ON THE SWEDISH INDUSTRIAL PRODUCTION

The most common way to evaluate methods for detection of turning points in business cycles is by using one set of data. We will now use that approach and use the actual data set on quarterly Swedish data on industrial production, presented in Section 3.1, to evaluate the three methods, PHM, MSR and SRlin. The evaluations made in Section 3.3 were made for a large number of realizations. Now we will use one specific realization, namely the real one. In the end of this section, we will discuss the relative merits of evaluation by simulation studies or by one example of real data.

According to official records (The National Institute of Economic Research (1992)), the peak occurs at time $t = 10$, implying that the time of change is $\tau=11$ (see Figure 27, where the evaluation period is presented in detail). This official time is not only based on the data on IP. Some other information might make this official time different than it should have been if only the IP data were used for this specific realization. The methods evaluated here only use the IP data. Figure 27 indicates that the turning point in the data is earlier than the official time. Thus, the methods could not be expected to be good at indicating the official time for this realization. All three methods give alarms earlier than the official times for this set of data.

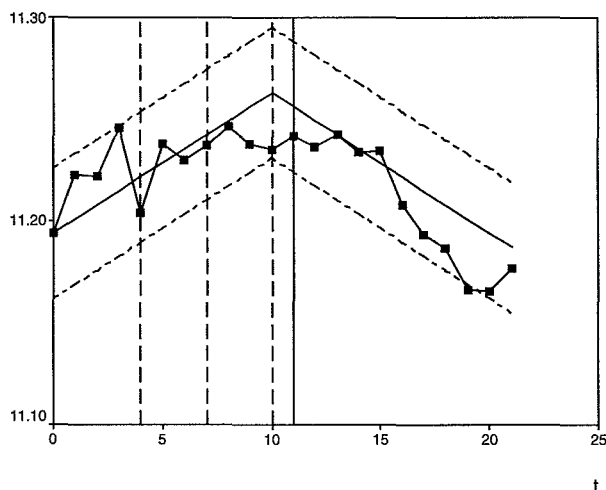


Figure 27: *Seasonally adjusted logarithm of the industrial production, quarterly data (1987Q2:1992Q2) for the evaluation period. The official time of change (11) is marked with a solid vertical line. The alarm times 4, 7 and 10 for the MSR, PHM, and SRLin respectively are marked with dashed vertical lines. The μ -curve for the piecewise linear model (14) with $\tau=11$ is marked with a solid curve. The 2.5th and 97.5th percentiles of values of the observations according to the model is marked as dotted curves.*

The piecewise linear model fits less well at the turning point as we have a plateau. McQueen and Thorley (1993), argues that it is reasonable that recessions tend to be preceded by plateaus. A plateau will result in a tendency to false alarms just before the turn. It can be discussed whether this is a drawback or not. An early indication of a coming recession is a plateau. In this light, the early alarms can be considered to be good, since they can be seen as warnings.

For PHM and SRLin it is assumed that all the parameters (slopes and standard deviations) are known or possible to estimate with great certainty. Here we use the observed data from the estimation period 1970Q1:1987Q1 and the estimation procedure described in Appendix 4 to estimate the parameters. The resulting signal-noise ratios are $\hat{\beta}_1/\hat{\sigma}_1 = 0.47$ for the expansion phase and $\hat{\beta}_2/\hat{\sigma}_2 = 0.40$ for the recession phase. These estimates are used for the PHM method when calculating the posterior probability for this realization. For the SRLin method we use a pooled (by their frequency) common estimate of the absolute value $\hat{\beta}/\hat{\sigma} = 0.41$ when calculating the likelihood ratio. The parameter values agree fairly well with the actual ones for this realization and with the one used in the simulation study, see Appendix 4.

A drawback with a simulation study based on a model is that the model might not be representative of the process we want to study. An actual data set is certainly representative of the specific time period and situation at hand. However, it might have stochastic deviation from the process of interest. The properties of the process are what is important for a method, when the method is intended to be used on future data. We will now study how often we should expect to get realizations as extreme as the one just described, if the model used in the simulations is true.

For the MSR method, the probability of such a large difference (for these specific times) as the dip from $t=3$ to $t=4$, which causes the alarm, is only 0.06.

The cumulative probability of an alarm for the situation of a turning point at $t = 11$ is presented in Figure 28.

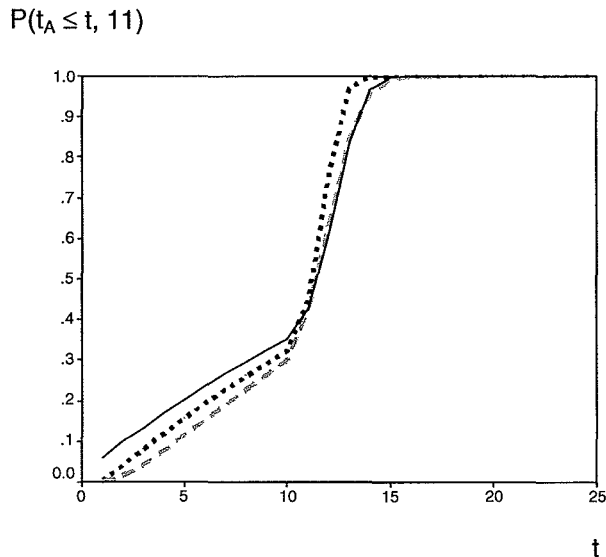


Figure 28: *The distribution of the time of an alarm, $\tau = 11$. PHM (—), MSR (---), SRlin (···).*

For PHM and SRlin we expect a false alarm rate of over 25% at their respective alarm times,

$$\text{PHM: } P(t_A \leq 7 \mid \tau = 11) = 0.27$$

$$\text{SRlin: } P(t_A \leq 10 \mid \tau = 11) = 0.32.$$

Evaluation of the properties of a method by one sample of real data is difficult, as demonstrated above. One difficulty is the definition of the turning point time with which to compare the results of the evaluation. Another difficulty is to know whether the turnout of the sample is a result that could be expected or if it is extreme. This difficulty has been avoided here. We can compare the results of the real-data evaluation with the results of the large-sample simulation study and thereby draw the conclusion that, in approximately one fourth of the cases, the SRlin and PHM methods will call an alarm at the same time as in this sample.

Evaluations by several real data sets (instead of just one) would decrease some of the stochastic variation in the measures of evaluation. However, if these analyses are not totally independent (for example if the same parameter estimates are used) then some of the stochastic components would keep their variance. Also, if the number of data sets is small, the essential disadvantages to the simulation study would remain.

5 SPECIAL DATA PROBLEMS

The discussion so far has been concerned with inferential problems relevant for all studies of detection of turning points in business cycles. However, for a specific application, there are many important problems before you have the ideal data sets to be analyzed. Some such problems will now be briefly mentioned.

5.1 Seasonal variation

Monthly or quarterly observations often contain seasonal variation, which could complicate the monitoring. The seasonal variation can be considerable, as is seen in Figure 2. If seasonality is neglected in the modeling and in the monitoring, it could lead to serious wrong conclusions. It is important that the structure of the original series is not disturbed by the seasonal adjustment. In a monitoring situation it is important that the time of the turning point is preserved after the adjustment.

The effect of using different filters is analyzed in Andersson and Bock (2001) and it is demonstrated that in a surveillance situation, using a Shewhart approach, the detection is delayed when a data transformation such as differentiating or moving average is used. The lowest probability for quick detection ($d < 4$) is caused by the moving average, whereas when more time is allowed for the detection to be considered "successful" ($d > 4$) the lowest probability is for the differentiated series. Most data-driven filters can have serious effects on the monotonicity. Thus, information from historical data or other prior knowledge, which makes the seasonal adjustment independent of the data to be analyzed, is very valuable.

5.2 Autocorrelation

Autoregressive models are often highly relevant when modeling economic time series. However, most suggested approaches for the detection of turns in business cycles assume that this is not a problem. Lahiri and Wang (1994) evaluate the performance of models of the form

$$y(t) = \mu(t) + \phi_1 \cdot \varepsilon(t-1) + \dots + \phi_4 \cdot \varepsilon(t-4) + \varepsilon(t),$$

where $\mu(t) = \{\mu_1, \mu_2\}$ and $\varepsilon(t) \sim N[0; \sigma]$,

and fitted models with autocorrelation of order $r = \{0, 1, 2, 3, 4\}$. They find that, using the same alarm limit in all five models, the introduction of autocorrelation in the errors leads to a smaller forecast error within phases (Brier's probability score) but increases the risk of wrong inference concerning turning points, as pointed out in Section 2.3.2. Ivanova et al. (2000) argue that the effect of the autoregressive parameters will largely be captured by the probabilities of remaining in the current state. The effect of autocorrelation can be dealt with by adjusting the alarm limit. Also, in the analysis in the earlier sections in this paper we have the assumptions for the model that the stochastic term ε is independent over time. The consequence of autocorrelation in the process is examined in the general theory of surveillance where also remedies are suggested. For a review, see Pettersson (1998) and Frisén (1999).

5.3 Adjusting for trend

Many macroeconomic variables can be characterized as cyclical movements around a trend. In order to distinguish the movements and make the time series stationary, it is sometimes necessary to adjust for the trend. In model (1), no separation between the trend and the cycle is made. This issue is treated differently in the PHM method and SRlin and MSR. In the PHM method differentiation is used, whereas in SRlin and MSR no adjustment is made.

The choice of a method for adjustment should depend on the assumptions regarding the trend-component. Whether the trend is assumed to have components that are deterministic, stochastic or both, has implications for the appropriate method,

see e.g. Enders (1995), p. 176. Adjusting for trend implies a data transformation, which may result in a distortion of the characteristics of the original series. Gordon (1997) studies the effect of trend removal for predictive densities of the US GDP and warns against using other information from the data than that which is directly associated with the business cycle turning points. Canova (1998) discusses trend removal and evaluates the effect using several different approaches, among them first order differentiating. One conclusion from the study is that linear trend removal does not result in turning points that correspond to the official turning point times of the National Bureau of Economic Research, USA. In another paper Canova (1999) points out that previous research has shown that the trend may interact with the cyclical component and is therefore difficult to isolate. The general conclusion is that statements concerning the turning points are not independent of the statistical assumptions needed to extract trends.

When analyzing short time series removing of the trend has less effect on the possibility to distinguish the turning points. The SRLin and MSR methods are applied to a part of the time series that contains one turning point at most. Thus, no attempt to separate the trend from the cycle is made.

5.4 Multivariate problems

By the common movement approach, a business cycle is characterized as the cyclical movement of many economical activities. This demonstrates that important information is contained in the relation between the turns of different indices. This information can be utilized, either by transforming the problem to a univariate one by using a composite index of leading indicators or by applying a multivariate method of surveillance.

Stock and Watson (1991) and Stock and Watson (1993) model the common movements of coincident variables as arising from an unobservable common factor that can be thought of as the overall state of the economy. The key element is the selection of variables and the estimation of the common factor. Leading indicators are added to the model to help predict future values of the common factor (overall state of economy). The probability that the economy will be in a recession six months hence is estimated.

Diebold and Rudebusch (1996) discuss the relation between the common movement approach and the regime approach and attempt to encompass both approaches by considering the common movements of coincident variables where the common factor depends on a hidden Markov chain with two states.

Kim and Nelson (1998) use the same approach as Diebold and Rudebusch (1996), but estimation is here made by Gibbs sampling in a Bayesian framework. They find that the ability to capture the common movement among several variables instead of just one, was the main cause of increase in forecast accuracy, whereas the prior assumptions concerning the transition probabilities had a minor influence. Hamilton and Perez-Quiros (1996) compare the accuracy (measured by the Brier probability score) in predicting the phases of U.S. real gross national product using univariate and bivariate linear models, where the latter included a composite leading index (CLI), and corresponding HMM. It is found that adding a CLI to the linear model results in the greatest increase in accuracy whereas using HMM makes no substantial increase in accuracy.

Birchenhall et al. (1999) exploit the feature of a business cycle, of common movements across variables, by extracting a business-cycle index from a vector of

time series. As in the works by Stock and Watson, the selection of variables is an important element.

Koskinen and Öller (1998) utilize multivariate information by monitoring a joint vector of leading indicators with a common time of turn. When used for turning point detection in the Swedish business cycle, the following three series are used: the Swedish Industrial Production, the Swedish Business Tendency Survey and Stockholm Stock Exchange Index. When the method is applied to the U.S. economy, the following two series are used: The first difference of Gross National Product and the Composite Index of Leading indicators.

Wessman (1998) demonstrates that the minimal sufficient alarm statistic, for changes in several variables with the same change point (or known time-lag), is univariate. The simulation study here has been made for the turning point of one leading index, possibly constructed as a function of many different indices. This is, in fact, the situation also for most of the earlier studies since a reduction to a univariate statistic is possible. However, procedures that are more efficient might be constructed by using the indices separately in the method of surveillance.

For reviews on multivariate surveillance, see Wessman (1999) and Frisé (1999). To use more of the theory of optimal multivariate surveillance for building a system for multivariate monitoring of business cycles might be a topic for future research.

6 DISCUSSION

Similarities between apparently different approaches are demonstrated. This might be a base for combining knowledge from several areas. Different approaches 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. As is demonstrated by the simulations, there is a high risk of false alarms soon after a turn if the knowledge of the type is not utilized. 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. When the limit for the posterior probability is fixed, 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 lower the false alarm rate, approaches such as smoothing are sometimes used.

The smoothing of the observations before the use of a method of monitoring will reduce the variation and hence reduce the false alarm probability. However, the smoothing also makes the turning points less pronounced and increases the expected delay of a desired alarm (for the same MRL^0). This is demonstrated by the simulation study. Also, the smoothing will introduce an autocorrelation. This will also introduce an extra dependency between successive decisions. Another way to reduce the effect of the variation of the observations, but keep control of the properties of the monitoring method, is to use a monitoring method that in itself is based on a moving average. The EWMA method uses exponential weights and is considered to have

good properties, even though it is not exactly optimal, as discussed by Frisé and Sonesson (2001).

As is seen in Figure 6, the PHM 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 long delays for changes that occur late (see Figure 9 and 10). However, the value of these early alarms can be questioned, as the predicted value is very low (see Figure 12). 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 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 task 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 MSR methods.

Parametric models contain information, which should be used whenever it is reliable. It is demonstrated that if the model for the process without turn is wrongly specified, this can have a great impact. A wrong specification of the process after a turn has less impact. One advantage of the non-parametric approach is that it works also when such reliable information on 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 MSR method only uses the monotonicity change and not the level. The wrong specifications of the slope used in the analysis are such that they could very well happen in practice just by stochastic variation. Thus, the very bad properties demonstrated for wrong specification of the slope give a warning. The safe way by the MSR method might be preferred.

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.

In Section 4 the three methods are used to analyze a period of the Swedish industrial production. A comparison between this evaluation of the methods and that of the simulation study is used to discuss the pros and cons of the two approaches.

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 (1999) 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.

REFERENCES

- Andersson, E. (2001) Monitoring cyclical processes. A non-parametric approach. To appear in *Journal of Applied Statistics*.
- Andersson, E. (2001) Turning point detection using non-parametric statistical surveillance. Evaluation of some influential factors. *Research report* 2001:3, Department of Statistics, Göteborg University, Sweden.
- Andersson, E. and Bock, D. (2001) On seasonal filters and monotonicity. *Research report* 2001:4, Department of Statistics, Göteborg University, Sweden.
- 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.
- Canova, F. (1998) Detrending and business cycle facts. *Journal of Monetary Economics*, **41**, 475-512.
- Canova, F. (1999) Does detrending matter for the determination of the reference cycle and the selection of turning points? *The Economic Journal*, **109**, 126-150.
- Chu, C.-S. J., Stinchcombe, M. and White, H. (1996) Monitoring structural change. *Econometrica*, **64**, 1045-1065.
- Delgado, M. A. and Hidalgo, J. (2000) Nonparametric inference on structural breaks. *Journal of Econometrics*, **96**, 113-144.
- Diebold, F. X., Lee, J.-H. and Weinbach, G. C. (1994) Regime Switching with Time-varying Transition Probabilities. In *Nonstationary Time Series Analysis* (Ed, Hargreaves, C.) Oxford University Press, Oxford, pp. 283-302.
- Diebold, F. X. and Rudebusch, G. D. (1989) Scoring the leading indicators. *Journal of Business*, **62**, 369-91.
- Diebold, F. X. and Rudebusch, G. D. (1996) Measuring business cycles: A modern perspective. *Review of Economics and Statistics*, **78**, 67-77.
- Enders, W. (1995) *Applied Econometric Time Series*.
- 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, A. J. (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* 1994:1 (Revised 2000), Department of Statistics, Göteborg University, Sweden.
- Frisén, M. (1999) Characterization of methods for surveillance by optimality. *Research report* 1999:13 (Revised 2000), Department of Statistics, Göteborg University, Sweden.
- Frisén, M. and de Maré, J. (1991) Optimal Surveillance. *Biometrika*, **78**, 271-80.
- Frisén, M. and Sonesson, C. (2001) Optimal surveillance by EWMA. *Research report* 2000:7, Department of Statistics, Göteborg University, Sweden.

- Frisén, M. and Wessman, P. (1999) Evaluations of likelihood ratio methods for surveillance. Differences and robustness. *Communications in Statistics. Simulations and Computations*, **28**, 597-622.
- Gan, F. (1993) An optimal-design of EWMA control charts based on median run-length. *Journal of Statistical Computation and Simulation*, **45**, 169-184.
- Gordon, S. (1997) Stochastic trends, deterministic trends, and business cycle turning points. *Journal of Applied Econometrics*, **12**, 411-434.
- Hackl, P. and Westlund, A. H. (1989) Statistical analysis of "Structural Change". *Empirical Economics*, **14**, 167-192.
- Hall, P., Marron, J. S. and Titterton, D. M. (1995) On partial local smoothing rules for curve estimation. *Biometrika*, **82**, 575-587.
- Hamilton, J. D. (1989) A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, **57**, 357-384.
- Hamilton, J. D. and Perez-Quiros, G. (1996) What Do the Leading Indicators Lead? *Journal of Business*, **69**, 27-49.
- Hawkins, D. L. (1992) Detecting shifts in functions of multivariate location and covariance parameters. *Journal of Statistical Planning and Inference*, **33**, 233-244.
- Hussey, R. (1992) Nonparametric Evidence on Asymmetry in Business Cycles Using Aggregate Employment Time Series. *Journal of Econometrics*, **51**, 217-231.
- Ivanova, D., Lahiri, K. and Seitz, F. (2000) Interest rate spreads as predictors of German inflation and business cycles. *International Journal of Forecasting*, **16**, 39-58.
- Jun, D. B. and Joo, Y. J. (1993) Predicting turning points in business cycles by detection of slope changes in the leading composite index. *Journal of Forecasting*, **12**, 197-213.
- Kim, C.-J. and Nelson, C. R. (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.
- Koskinen, L. and Öller, L.-E. (1998) A Classifying Procedure for Signalling Turning Points. *Working Paper*, National institute of economic research, Stockholm, Sweden.
- Lahiri, K. and Moore, G. (1991) *Leading economic indicators: New approaches and forecasting record*, Cambridge University Press, Cambridge .
- Lahiri, K. and Wang, J. G. (1994) Predicting cyclical turning points with a leading index in a Markov Switching model. *Journal of Forecasting*, **13**, 245-263.
- Lai, T. L. (1995) Sequential Change-point Detection in Quality-Control and Dynamical-Systems. *Journal of the Royal Statistical Society B*, **57**, 613-658.
- Layton, A. P. (1996) Dating and predicting phase changes in the U.S. business cycle. *International Journal of Forecasting*, **12**, 417-428.
- Li, D. T. and Dorfman, J. H. (1996) Predicting turning points through the integration of multiple models. *Journal of Business and Economic Statistics*, **14**, 421-428.
- McQueen, G. and 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., 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.
- Pettersson, M. (1998) Evaluation of some methods for statistical surveillance of an autoregressive process. *Research report* 1998:4, Department of Statistics, Göteborg University, Sweden.
- Roberts, S. W. (1966) A Comparison of some Control Chart Procedures. *Technometrics*, **8**, 411-430.
- Robertson, T., Wright, F. T. and Dykstra, R. L. (1988) *Order Restricted Statistical Inference*, John Wiley & Sons Ltd.
- Shiryayev, A. N. (1963) On optimum methods in quickest detection problems. *Theory of Probability and its Applications.*, **8**, 22-46.
- Srivastava, M. S. and Wu, Y. (1993) Comparison of EWMA, CUSUM and Shiryayev-Roberts Procedures for Detecting a Shift in the Mean. *Annals of Statistics*, **21**, 645-670.
- Stock, J. H. and Watson, M. W. (1991) A Probability Model of the Coincident Economic Indicators. In *Leading Economic Indicators: New Approaches and Forecasting Records* (Eds, Lahiri, K. and Moore, G. H.) Cambridge University Press, Cambridge, pp. 63-89.
- Stock, J. H. and Watson, M. W. (1993) A Procedure for Predicting Recessions With Leading Indicators: Econometric Issues and Recent Experience. In *Business Cycles, Indicators and Forecasting*, Vol. 28 (Eds, Stock, J. H. and Watson, M. W.) University of Chicago Press, Chicago.
- Wessman, P. (1998) Some Principles for surveillance adopted for multivariate processes with a common change point. *Communications in Statistics. Theory and Methods*, **27**, 1143-1161.
- Wessman, P. (1999) Studies on the surveillance of univariate and multivariate processes. PhD Thesis at the Department of Statistics, Göteborg University, Sweden, pp. 103.
- Westlund, A. H. and Zackrisson, U. (1986) On the prediction of structurally varying systems. *Technological Forecasting and Social Change*, **30**, 63-72.
- Wetherhill, G. B. and Brown, D. W. (1991) *Statistical Process Control: Theory and Practice*, Chapman and Hill, London.
- Zarnowitz, V. and Moore, G. H. (1982) Sequential signals of recessions and recovery. *Journal of Business*, **55**, 57-85.
- Zellner, A., Hong, C. and 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.
- Öller, L.-E. (1986) A note on exponentially smoothed seasonal differences. *Journal of Business and Economic Statistics*, **4**, 485-489.
- Öller, L.-E. (1987) Stabiliseringspolitik och prognoser. *Ekonomiska samfundets tidskrift*, **40**, 23-27.
- Öller, L.-E. and Tallbom, C. (1994) Hybrid indicators for the Swedish economy based on noisy statistical data and the business tendency survey. *Working Paper*, National institute of economic research, Stockholm, Sweden.

Appendix 1. The optimal alarm rule for linear functions, LRLin

The alarm rule, for discriminating between events C and D , which fulfills several optimality conditions according to Frisé and de Maré (1991) is

$$x_s : \frac{f(x_s|C)}{f(x_s|D)} > \frac{k \cdot P(D)}{(1-k) \cdot P(C)} \text{ or equivalently } P(C|x_s) > k,$$

where k is a constant.

When C is a composite event, the alarm rule is

$$\frac{\sum_{j=1}^s v_j \cdot f(x_s|C_j)}{f(x_s|D)} > \frac{k \cdot P(D)}{(1-k) \cdot P(C)},$$

where $v_j = P(C_j) / P(C)$.

The events $C = \{C_1, C_2, \dots, C_s\}$ and D are here expressed in terms of μ -vectors and by $C_j = \{\tau \leq j\}$, $C = \{\tau \leq s\}$ and $D = \{\tau > s\}$ the alarm rule is

$$\frac{\sum_{j=1}^s v_j \cdot f(x_s|\mu = \mu^{C_j})}{f(x_s|\mu = \mu^D)} > k_s,$$

where $k_s = \frac{k \cdot P(\tau > s)}{(1-k) \cdot P(\tau \leq s)}$.

Under assumption of normal distribution the alarm rule is

$$\sum_{j=1}^s v_j \cdot \exp \left[\left(\frac{1}{2\sigma^2} \left(\sum_{t=1}^s (x(t) - \mu^D(t))^2 - \sum_{t=1}^s (x(t) - \mu^{C_j}(t))^2 \right) \right) \right] > k_s.$$

Since $\mu^{C_j}(t) = \mu^D(t)$ for $t < \tau$, $j = \{1, 2, \dots, s\}$, the alarm rule can be written as

$$\sum_{j=1}^s v_j \cdot \exp \left[\left(\frac{1}{2\sigma^2} \left(\sum_{t=j}^s (x(t) - \mu^D(t))^2 - \sum_{t=j}^s (x(t) - \mu^{C_j}(t))^2 \right) \right) \right] > k_s.$$

With

$$\mu^D = \beta_0 + \beta_1 \cdot t$$

and

$$\mu^{C_j} = \beta_0 + \beta_1 \cdot (j-1) - \delta_1 \cdot (t-j+1)$$

we have the alarm rule

$$\text{LRlin}(s) > \frac{k \cdot P(\tau > s)}{(1-k) \cdot P(\tau \leq s)},$$

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

$$\begin{aligned} & \sum_{j=1}^s v_j \cdot \exp \left[\left(\frac{1}{2\sigma^2} \right) \left(2(-\delta_1 - \beta_1) \cdot \sum_{u=j}^s (x(u) \cdot u) + 4\delta_1 \cdot \sum_{u=j}^s (x(u) \cdot (j-1)) \right) \right] * \\ & \exp \left[\left(\frac{1}{2\sigma^2} \right) \left((\beta_1^2 - \delta_1^2) \cdot \sum_{u=j}^s u^2 + 4\delta_1^2 \cdot (j-1) \cdot \left(\sum_{u=j}^s (u-j+1) \right) \right) \right] * \\ & \exp \left[\left(\frac{1}{2\sigma^2} \right) \left(2\beta_0 \cdot (\beta_1 + \delta_1) \cdot \sum_{u=j}^s u - 4\beta_0 \delta_1 \cdot \sum_{u=j}^s (j-1) \right) \right]. \end{aligned}$$

For the symmetric case, $\beta_1 = \delta_1$, we have

$\text{LRlin}(s) =$

$$\begin{aligned} & \sum_{j=1}^s v_j \cdot \exp \left[\left(\frac{1}{2\sigma^2} \right) \left(4\beta_1 \cdot \sum_{u=j}^s (x(u) \cdot (j-1-u)) \right) \right] * \\ & \exp \left[\left(\frac{1}{2\sigma^2} \right) \left(4 \cdot \beta_1^2 \cdot (j-1) + 4 \cdot \beta_0 \cdot \beta_1 \cdot \sum_{u=j}^s (u-j+1) \right) \right]. \end{aligned}$$

Appendix 2. Knowledge of the type of the next turn

We will here illustrate the effect of knowledge of the next type of turn by giving details for the case of $s=3$. We start by expressing the alarm rule (11) in Section 2.5,

$$P(C(3) | x_3) > k,$$

as a ratio of two conditional probabilities

$$\frac{f(x_3 | C(3)) \cdot P(C(3))}{f(x_3 | D(3)) \cdot P(D(3))} > \frac{k}{1-k},$$

which is true as soon as $D(3)$ is the complement to $C(3)$.

In the following, the variable $J(t)$ denotes the state at time t , so that

$$J(t) = \begin{cases} 1 \\ 2, \end{cases}$$

where state 1 is the expansion state and state 2 is the recession state.

In the HMM approach, the event $C_{\text{HMM}}(3)$ is consistent with the following 2^{3-1} possible combinations for $J(1), J(2), J(3)$:

$$\begin{aligned} \{J(1)=2, J(2)=2, J(3)=2\} &\text{ denoted } C^{\text{I}}, \\ \{J(1)=1, J(2)=2, J(3)=2\} &\text{ denoted } C^{\text{II}}, \\ \{J(1)=1, J(2)=1, J(3)=2\} &\text{ denoted } C^{\text{III}}, \\ \{J(1)=2, J(2)=1, J(3)=2\} &\text{ denoted } C^{\text{IV}}, \end{aligned}$$

and the event $D_{\text{HMM}}(3)$ is consistent with the following 2^{3-1} combinations:

$$\begin{aligned} \{J(1)=1, J(2)=1, J(3)=1\} &\text{ denoted } D^{\text{I}}, \\ \{J(1)=2, J(2)=1, J(3)=1\} &\text{ denoted } D^{\text{II}}, \\ \{J(1)=2, J(2)=2, J(3)=1\} &\text{ denoted } D^{\text{III}}, \\ \{J(1)=1, J(2)=2, J(3)=1\} &\text{ denoted } D^{\text{IV}}. \end{aligned}$$

An illustration of $C^{\text{I}}, \dots, C^{\text{IV}}$ and $D^{\text{I}}, \dots, D^{\text{IV}}$ is given in Figure 1A below.

The probabilities $P(C^i)$ and $P(D^i)$, $i = \{\text{I, II, III, IV}\}$, are functions of the transition probabilities. For example

$$\begin{aligned} P(C^{\text{I}}) &= \\ P(J(1)=2) \cdot P(J(2)=2 | J(1)=2) \cdot P(J(3)=2 | J(2)=2) &= \\ P(J(1)=2) \cdot p_{22} \cdot p_{22}. \end{aligned}$$

Analogously we have

$$\begin{aligned} P(C^{\text{II}}) &= P(J(1)=1) \cdot p_{12} \cdot p_{22}, \\ P(C^{\text{III}}) &= P(J(1)=1) \cdot p_{11} \cdot p_{12}, \\ P(C^{\text{IV}}) &= P(J(1)=2) \cdot p_{21} \cdot p_{12}, \end{aligned}$$

and

$$\begin{aligned} P(D^{\text{I}}) &= P(J(1)=1) \cdot p_{11} \cdot p_{11}, \\ P(D^{\text{II}}) &= P(J(1)=2) \cdot p_{21} \cdot p_{11}, \\ P(D^{\text{III}}) &= P(J(1)=2) \cdot p_{22} \cdot p_{21}, \\ P(D^{\text{IV}}) &= P(J(1)=1) \cdot p_{12} \cdot p_{21}. \end{aligned}$$

If we know that the next turn is a peak, then we have no transitions from recession to expansion, that is $p_{21}=0$, and hence $p_{22}=1$. Then,

$$P(C^{IV}) = P(D^{II}) = P(D^{III}) = P(D^{IV}) = 0,$$

and the PHM alarm rule given above is reduced to

$$\frac{f(x_3|C^I) \cdot P(C^I) + f(x_3|C^{II}) \cdot P(C^{II}) + f(x_3|C^{III}) \cdot P(C^{III})}{f(x_3|D^I) \cdot P(D^I)} > \frac{k}{1-k}.$$

This is equivalent to the alarm rule in the surveillance approach (with knowledge of type of next turn),

$$\frac{\sum_{j=1}^3 f(x_3|\mu = \mu^{C_j}) \cdot P(C_j)}{f(x_3|\mu = \mu^D) \cdot P(D)} > \frac{k}{(1-k)}$$

since

$$C^I = \{J(1)=2, J(2)=2, J(3)=2\} = C_1 = \{\tau=1\},$$

$$C^{II} = \{J(1)=1, J(2)=2, J(3)=2\} = C_2 = \{\tau=2\},$$

$$C^{III} = \{J(1)=1, J(2)=1, J(3)=2\} = C_3 = \{\tau=3\},$$

$$D^I = \{J(1)=1, J(2)=1, J(3)=1\} = D = \{\tau > 3\}.$$

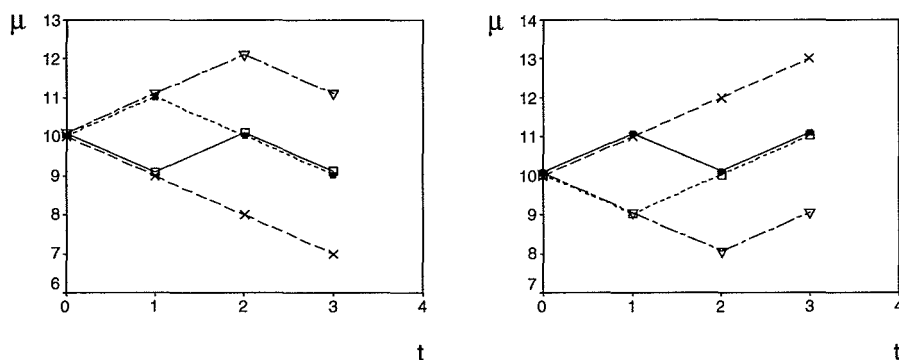


Figure 1A: Illustration of the different possible paths of μ -vector for $s = 3$. To the left are examples consistent with C_{HMM} (C^I (---x), C^{II} (···■), C^{III} (-·-▽), C^{IV} (—□)). To the right are paths consistent with D_{HMM} (D^I (---x), D^{II} (···□), D^{III} (-·-▽), D^{IV} (—■)).

Appendix 3. Regression functions used to simulate the estimation period.

For each expansion and recession phase, a second degree polynomial regression function with seasonal dummy variables is fitted to data on the $\ln(IP)$. After that, the polynomials are intercept-adjusted, not to expose any jumps.

Results are given in Table 1A for each of the four recession phases and each of the three expansion phases, according to the model (without seasonal variation)

$$y_{ij}(t) = \theta_{0ij} + \theta_{1ij} \cdot t + \theta_{2ij} \cdot t^2 + \varepsilon_{ij}(t),$$

where $i = \{\text{expansion}\}$ or $\{\text{recession}\}$

and $j = \{1, 2, 3, 4\}$.

For the dating of recession and expansion phases, official records are used (National Institute of Economic Research (1992)).

Table 1A

<i>i</i>	<i>j</i>	θ_0	θ_1	θ_2	$sd[\varepsilon(t)]$
Exp	1	10.707	0.02315	-0.000172	0.003618
	2	11.056	-0.0188	0.0004724	0.0235
	3	12.678	-0.0753	0.0008423	0.0131
Rec	1	10.920	0.008721	-0.000938	0.00667
	2	10.550	0.04589	-0.00103	0.0183
	3	8.426	0.117	-0.00132	0.01255
	4	11.615	-0.0163	0.0001473	0.01983

Appendix 4. Estimation of slopes during the estimation period

For the PHM method, the regression

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

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

is estimated using data of the estimation period. The estimation procedure is described below.

Each time point is classified as being either expansion or recession by the same procedure as in Koskinen and Öller (1998) where the following definition of a turning point is used:

The seasonally differentiated series has kept the same sign for at least two consecutive time points when it changes sign. If the new sign is kept during at least the next time point, then a turning point is said to have occurred.

The method for estimating $\{\beta_1, \beta_2, \sigma_1^2, \sigma_2^2\}$ from quarterly data, described below, is one component in the estimation procedure used by Koskinen and Öller (1998):

$$\hat{\beta}_1 = \frac{\bar{d}_1}{4} = \frac{1}{4 \cdot n_1} \sum_{i=1}^{n_1} d_{1i},$$

$$\hat{\beta}_2 = \frac{\bar{d}_2}{4} = \frac{1}{4 \cdot n_2} \sum_{i=1}^{n_2} d_{2i}$$

and

$$\hat{\sigma}_1^2 = \frac{1}{2 \cdot (n_1 - 1)} \sum_{i=1}^{n_1} (d_{1i} - \bar{d}_1)^2,$$

$$\hat{\sigma}_2^2 = \frac{1}{2 \cdot (n_2 - 1)} \sum_{i=1}^{n_2} (d_{2i} - \bar{d}_2)^2,$$

where $n_1 = \#$ time point classified as expansions,

and $n_2 = \#$ time point classified as recessions,

and d_1 and d_2 is the seasonally differentiated series classified as belonging to expansion and recession, respectively.

The resulting $\hat{\beta}_1$ and $\hat{\beta}_2$ with standard errors are presented below along with $\hat{\sigma}_1$ and $\hat{\sigma}_2$.

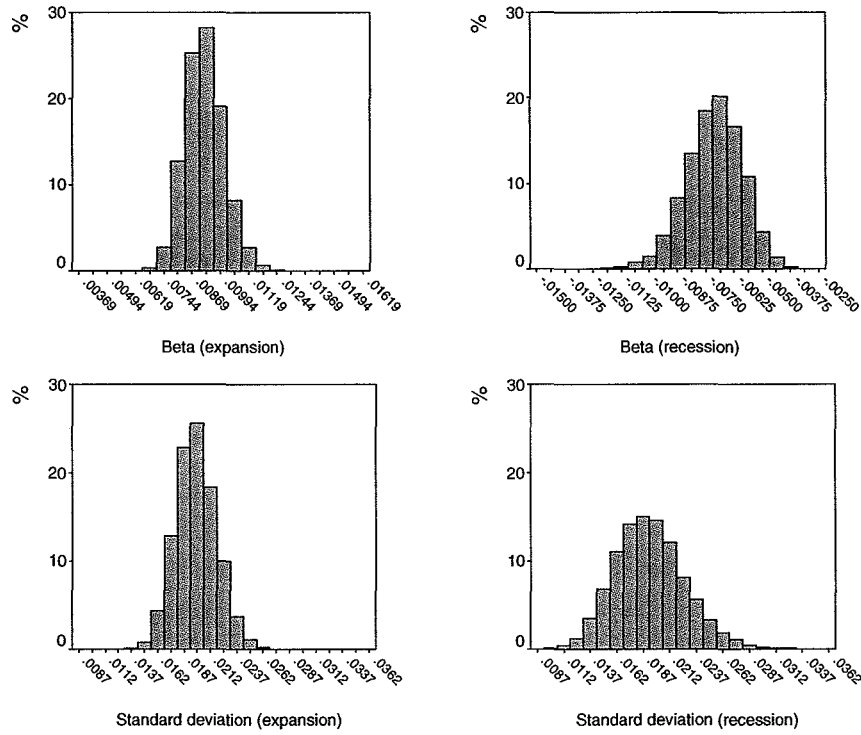


Figure 2A: The frequencies for the estimates of the parameters β_1 , β_2 , σ_1 , σ_2 for the estimation period used in the simulation study (see Section 3.2.1).

The mean and standard deviation for each parameter is given below.

Table 2A: Mean and standard deviation of parameters

	Expansion		Recession	
	$\hat{\beta}_1$	$\hat{\sigma}_1$	$\hat{\beta}_2$	$\hat{\sigma}_2$
Mean	0.0092	0.020	-0.0071	0.019
Standard deviation	0.0009	0.0019	0.0012	0.0033

Research Report

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