WORKING PAPER No. 59
April, 1998

A HIDDEN MARKOV MODEL AS A DYNAMIC BAYESIAN CLASSIFIER, WITH AN APPLICATION TO FORECASTING BUSINESS-CYCLE TURNING POINTS

LASSE KOSKINEN
LARS-ERIK ÖLLER

Svensk resumé på sid 23.

A Hidden Markov Model as a Dynamic Bayesian Classifier, With an Application to Forecasting Business-Cycle Turning Points

Lasse Koskinen and Lars-Erik Öller *

* Lasse Koskinen is Statistician, National Institute of Economic Research, P.O. Box 3116, S-103 62, Stockholm, Sweden (E-mail: lasse.koskinen@konj.se); Lars-Erik Öller is Director of Research, National Institute of Economic Research, P.O.Box 3116, S-103 62, Stockholm, Sweden, and Professor, Stockholm School of Economics, P.O.Box 6501, S-113 83 Stockholm, Sweden (E-mail:lars_erik.oller@konj.se).

Abstract. We introduce a method for dynamic classification of vector time series data into different regimes. A hidden Markov regime-switching model is used in classification. Past regimes are determined in advance and characterized by first and second moments of the observation vector. In estimation and model selection, instead of the maximun likelihood principle, we use Brier's probability score making it possible to perform feature extraction, eg. noise-removing filtering. When calibrated to the forecast horizon, the method provides a simple and computationally efficient way to utilize leading information in forecasting regimes in time series. The method is applied on forecasting turning points of Sweden's industrial production, where the Stock Market Index and a Business Tendency Survey series together express expectations, providing leading information. The method is also tested on forecasting the business cycle of the US, using GDP and the Department of Commerce Composite Index of Leading Indicators.

Key words. Empirical Bayesian, Expectation, Leading Indicator, Pattern Recognition, Probability Forecast, Regime-Switching Model.

1. INTRODUCTION

In this paper, turning point prediction is interpreted as a dynamic classification problem and a new classifier is introduced. The prediction of turning points is both conceptually and methodologically a different problem from making point forecasts of a time series. Generally, it is difficult and requires large data sets to forecast every value of a time serie, whereas turning points may be easier to forecast when the method is focussed on that problem. For economic agents, getting reliable directional forecasts is often more important than general numerical forecast accuracy, cf. Leitch and Tanner (1995). Methods currently in use keep transmitting wrong turning point signals, and the field is wide open to new techniques, cf. Funke (1997), and Stock and Watson (1993).

The present paper develops a probabilistic classification method that is particularly suitable for producing probability forecasts of turning points, where leading information is taken from other series. Sometimes, only a part of the external information can help in forecasting turning points. In Öller and Tallbom (1996) low frequency or high amplitude data are considered. Here leading information is extracted from data patterns as they appear in first and second moments.

When a continuous random variable is forecasted it is common practice to provide both a point forecast and its confidence interval. For a dichotomous random variable, e.g. a turning point indicator, some other concept has to be chosen, such as the probability of an occurrence and risk bounds for false inference. A number of probability methods for turning point forecasting have been suggested in the literature. Most of them are based either on Neftci's (1982) or Hamilton's (1989) business cycle models. Neftci's model uses sequential probabilities. Hamilton's probabilistic business-cycle model adopts the *hidden Markov regime-switching model* (HMM) from Lindgren (1978). The model has been further generalized in McCulloch and Tsay (1994).

Recently, Hamilton's model has been applied in constructing leading indicators, cf. Layton (1996), Lahiri and Wang (1994,1996), Hamilton and Lin (1996), and Hamilton and Perez-Quiros (1996). Bayesian inference and prediction using HMM have been developed in McCulloch and Tsay (1993). The present paper has Lindgren's HMM as its data generator. Our method differs from earlier attempts in that we apply HMM only as a probabilistic classifier; the classes (regimes) being identified in advance as either: expansion or recession. For an interesting account of how to do this discrimiantion, cf. McNees (1991) and Boldin (1994). This is not merely a way of interpreting HMM; it implies other model building procedures, requiring observed turning points to be specified in advance, not estimated. Since in a practical forecasting situation this information is available, why not use it! Also, there are several ways to define a turning point, and it may be an advantage to give the forecaster a chance to choose an appropriate definition.

Pattern recognition is done in three stages: feature extraction, classification and evaluation, cf. Fukanaga (1990), and in a time series context, Shumway (1982). Classification into recession or no recession is obscured by noise in economic data, cf. Morgenstern (1965). Transformations, such as smoothing, are used to facilitate discrimination and this is called "feature extraction", which in the present paper's applications simply means filtering by an exponentially weighted moving average. Our general approach is close to Artis et al. (1996), where turning point prediction means pattern recognition, applying Neftci's model for making probability forecasts; here we use HMM as a probabilistic classifier, more precisely, a *Markov-Bayesian classifier* (MBC). We propose Brier's probability score, ie. least squares, as an estimation criterion, because when taking the pattern recognition approach, maximum likelihood is inappropriate.

The classification method produces a leading probability indicator for turning points in industrial production in Sweden, where a rule of thumb is used for dating the business cycle. Some monetary and financial leading indicators have recently been presented in connection with HMM. Stock and Watson (1989) recommend as a leading series the interest rate spread (the same in Lahiri and Wang (1996)), and Hamilton and Lin (1996) pick stock market volatility. Here the leading series are the *Stockholm Stock Market Index* (SSMI) and answers to a *Business Tendency Survey* (BTS) question. We also check our method on forecasting US turning points, as determined by NBER, using GDP and the *Department of Commerce Composite Index of Leading Indicators* (CLI).

The paper is organized as follows. In the next section the MBC is introduced. The third section proposes a turning point forecasting method based on MBC. The fourth section is devoted to empirical applications. Some conclusions are drawn in the final section.

2. A MARKOV-BAYESIAN CLASSIFIER

We begin this section by describing some concepts for static classification (cf. Fukanaga (1990)), subsequently extending the method to time series. The result is a probability classifier whose dynamics is governed by a hidden Markov chain.

2.1. A Mahalanobis Distance Classifier

We consider two classes (a generalization to an arbitrary number of classes is straight-forward.) $i = \{1, 2\}$ and a vector $y \in \mathbf{R}^n$ of data to be allocated into either of these classes. Formally, there is a pair $Z = (Y, J)$, where Y is a random vector and $J : \Omega \to \{1, 2\}$ is a random variable that assigns class information to Y, Ω being the sample space. We observe only Y, whereas J is hidden. Thus, one needs a rule (function) $g : \mathbf{R}^n \to \{1, 2\}$, that as accurately as possible assigns an observed vector $y, (Y = y)$, to the right class. We assume at this stage, that both the means and the covariance matrices, μ_i and V_i, of the classes $i = \{1, 2\}$ are known.

The *Mahalanobis's distance*

$$D_V^2(\mu, y) = (y - \mu)' V^{-1} (y - \mu) \tag{1}$$

is a frequently used measure of how far a random vector is located from the mean of its distribution. A reasonable classification rule is to assign y to the class that minimizes the observed Mahalanobis distance. In other words, allocate y to class r, $g(y) = r$, if

$$(y - \mu_r)' V_r^{-1} (y - \mu_r) = \min_i \{(y - \mu_i)' V_i^{-1} (y - \mu_i)\}, \; i = \{1, 2\}. \tag{2}$$

2.2 A Gaussian Classifier

If the observations can be assumed to have a multivariate Gaussian distribution, $y \sim N(\mu_i, V_i)$ for $J = i$, the maximum likelihood (ML) rule is close to the Mahalanobis distance rule. Denoting $\theta = \{\mu_1, V_1, \mu_2, V_2\}$, the *Gaussian classifier* allocates y to class r, $g(y) = r$, if

$$log|V_r| + (y - \mu_r)' V_r^{-1} (y - \mu_r) = \min_i \{log|V_i| + (y - \mu_i)' V_i^{-1} (y - \mu_i)\}. \tag{3}$$

The Mahalanobis distance (1) just defines a metric, while (3), including $log|V_i|$, is an optimal classification device for Gaussian distributions.

2.3 A Static Probability Classifier

Our next task is to derive a classifier that assigns a probability to the event that an observation is from class i. Denoting the prior probabilities $P\{J = i\}$ by $p(i)$, class (posterior) probabilities for observed y can be calculated by *Bayes' rule:*

$$P\{J = i \mid Y = y\} = \frac{p(i) \times f(y \mid J = i)}{p(1) \times f(y \mid J = 1) + p(2) \times f(y \mid J = 2)}, \tag{4}$$

where $f(y_t \mid J = i)$ is the density function of class i. Further, when $P\{J = i \mid Y = y\}$ are known, the *Bayesian classifier* allocates y to r, $g(y) = r$, if

$$P\{J = r \mid Y = y\} \geq \frac{1}{2}. \tag{5}$$

Class probabilities express the uncertainty: the closer the probabability estimate is to one or to zero, the less uncertainty there is in a decision, where 1/2 is the threshold in (5). The *Classification error probability* $R\{Y\}$ when allocating Y is

$$R\{Y\} = P\{g(Y) \neq J\}. \tag{6}$$

2.4 A Dynamic Probability Classifier: MBC

The classifier defined in (5) could be called static. However, our aim is to form a dynamic probability classifier that also describes the dependence between observations when dealing with time series data, $Z = Z_t = (Y_t, J_t)$. The normality assumption of classes now takes the form:

$$Y_t \sim N(\mu_{J_t}, V_{J_t}). \tag{7}$$

In order to model time dependence in a mathematically tractable way, we postulate that J_t can be described as a homogeneous Markov chain, where the data generating process has two hidden classes. For each class the likelihood of various observations is either of the two multinormal densities given in (7). A Markov chain generates switching between classes. When in class i, the process is said to be working in *regime* i. For each $i, j \in \{1, 2\}$ we assume a constant probability $p_{ij} = Prob\{J_t = i \mid J_{t-1} = j\}$ that regime j will be followed by regime i. These *Markov probabilities* are collected into a Markov matrix $\mathbf{P} = (p_{ij})$. The model is now defined by regime distributions (7) and the Markov matrix $\mathbf{P}$. The Markov probabilities are p_{11}, p_{22} $(p_{11} + p_{21} = 1, p_{22} + p_{12} = 1)$, and the complete parameter set is $\Theta = \{\mu_1, \mu_2, V_1, V_2, p_{11}, p_{22}\}$. When class priors at $t-1$, $p_i(t-1)$, are known, the regime (posterior) probabilities for given y_t are

$$P\{J_t = i \mid Y_t = y_t\}$$
$$= \frac{(p_{1i} \times p_1(t-1) + p_{2i} \times p_2(t-1)) \times f(y \mid J_t = i)}{\sum_{j=1}^{2} (p_{1j} \times p_1(t-1) + p_{2j} \times p_2(t-1)) \times f(y \mid J_t = j)}. \tag{8}$$

The formulae (7-8) and (5) define the *Markov-Bayesian Classifier* (MBC). Conditional densities $f(y_t \mid J_t)$ in (8) can be replaced by the inverses of the Mahalanobis distances (1) when the regime distributions are strongly non-normal. Exact classification error probabilities (6) cannot be derived analytically. Upper and lower bounds are given by

$$\frac{1}{2}\left(1 - \sqrt{1 - 4p_1 \times (1 - p_1)\, e^{-2B^2}}\right) \le R\{Y_t\} \le \sqrt{p_1 \times (1 - p_1)}\, e^{-B^2}, \tag{9}$$

based on the *Bhattacharyya distance* B^2 between classes 1 and 2

$$B^2 = \frac{1}{8}D_V^2(\mu_1, \mu_2) + \frac{1}{2}log\Big(\frac{|V|}{\sqrt{|V_1| \times |V_2|}}\Big), \tag{10}$$

where $V = \Big(V_1 + V_2\Big)/2$ and D_V^2 is the Mahalanobis distance (1).

2.5 The Estimation of MBC Regime Probabilities

The introduced model is a special case of HMM. Y_t is an observed time series depending on an unobserved Markov chain J_t. When the parameter set Θ is known and $\{y_0, \ldots, y_T\}$ is an observed time series, whose subsample is denoted by $\mathbf{Y}_t = \{y_0, \ldots, y_t\}$, $0 \leq t \leq T$, it is possible to estimate the classifying probability, $P\{J_t = i \mid \mathbf{Y}_t, \Theta\}$, for each observed vector y_t. Hamilton (1994, page 692) presents an estimation algorithm in a recursive form, utilizing Bayes' rule and the Markov property. Collect the posterior probability estimates $\hat{P}\{J_t = i \mid \mathbf{Y}_t, \Theta\}$ of (8) and related Markov forecasts for priors $P\{J_{t+1} = i \mid \mathbf{Y}_t, \Theta\}$ into 2×1 vectors. Given starting values of probabilities at time zero, an estimate for the classifying probability of y_t can be found by iterating the following pair of equations:

$$\begin{pmatrix} \hat{P}\{J_t = 1 \mid \mathbf{Y}_t, \Theta\} \\ \hat{P}\{J_t = 2 \mid \mathbf{Y}_t, \Theta\} \end{pmatrix} = \frac{\begin{pmatrix} \hat{P}\{J_t = 1 \mid \mathbf{Y}_{t-1}, \Theta\} \times \; f(y_t \mid J_t = 1, \Theta) \\ \hat{P}\{J_t = 2 \mid \mathbf{Y}_{t-1}, \Theta\} \times \; f(y_t \mid J_t = 2, \Theta) \end{pmatrix}}{f(y_t \mid \Theta)}, \tag{11}$$

and

$$\begin{pmatrix} \hat{P}\{J_{t+1} = 1 \mid \mathbf{Y}_t, \Theta\} \\ \hat{P}\{J_{t+1} = 2 \mid \mathbf{Y}_t, \Theta\} \end{pmatrix} = \begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix} \begin{pmatrix} \hat{P}\{J_t = 1 \mid \mathbf{Y}_t, \Theta\} \\ \hat{P}\{J_t = 2 \mid \mathbf{Y}_t, \Theta\} \end{pmatrix}, \tag{12}$$

where

$$\begin{aligned} &f(y_t \mid \Theta) \\ &= \hat{P}\{J_t = 1|\mathbf{Y}_{t-1}, \Theta\} \times f(y_t|J_t = 1, \Theta) + \hat{P}\{J_t = 2|\mathbf{Y}_{t-1}, \Theta\}) \times f(y_t|J_t = 2, \Theta). \end{aligned} \tag{13}$$

In this paper we set a neutral starting value, 1/2. We return to the problem of estimating Θ in the next section.

3. CONSTRUCTING A TURNING POINT INDICATOR

In this section, we demonstrate how MBC, can be used for turning point forecasting, a natural application for two reasons. Firstly, Diebold and Rudebush (1990) concluded that a simple Markov chain provides a reasonable description of the traditional NBER business cycle dates. Secondly, a pattern recognition approach supports the use of vector series whose components carry leading information. We propose an iteration of feature extraction, classification and evaluation.

Assume that we are standing at time t and that we know exactly in which regime we have been at every point in the past up to t. It is then possible to estimate the means and variances for each of the two regimes. When we now try to look past t into the future, the situation turns into a genuine HMM problem, where, however, the static regime parameters $\{\mu_i, V_i\}$ have been estimated in advance. Another feature that distances our approach from how HMM has been used, is the way the model is estimated. Usually a maximum likelihood approach is taken also for forecasting. In the spirit of Geisser and Eddy (1979), we adopt a predictive approach to model specification, basing parameter estimation and model selection on a probability score, minimizing the turning point forecast error. This provides a possibility for feature extraction, eg. removing noise. The computational complexity of the method depends on the class of filters studied (if any) and the optimization algorithm used.

Let J_t denote the regime series. Our task is to predict $J_{t+l},\ l > 0$ at t, applying MBC on series y_t which carries leading information on J_{t+l} (If $l = 0$ the indicator is said to be contemporaneous). We have the same model as in the previous section, but observation Y_t is leading and comes from regime J_{t+l} ie. $Z_t = (Y_t, J_{t+l})$. Let an observation set of the series Y_t, containing leading information, be $\{y_0, y_1, \ldots, y_T\}$. For estimating regime parameters we divide the dates $I = \{1, \ldots, T\}$ into two sets:

$$I(i) = \{t \in I \mid J_{t+l} = i\}. \qquad (14)$$

The number of elements in $I(i)$ is denoted by T_i.

Hamilton (1989) and Hamilton and Perez-Quiros (1996) have proposed the use of (12) to forecast regimes. However, the transition Matrix P is constant and works as a linear operator. Hence, it does not provide optimal information on turning points. For our method, it is crucial that one calibrates the model to horizon l and *elicits information* from y_t on J_{t+l}.

We propose that model estimation follows a three stage procedure, where stages are iterated with different feature extraction filters, until an optimal model is found.

Step I. *Feature extraction.* Apply a causal filter F, chosen so as to facilitate classification and to provide correct reactions at turning points. We denote $\tilde{y}_t = F(y_t)$.

Step II. *Estimation of regime parameters.* Estimate:

$$\hat{\mu}_i = \frac{1}{T_i} \sum_{t \in I(i)} \tilde{y}_t \tag{15}$$

and

$$\hat{V}_i = \frac{1}{T_i} \sum_{t \in I(i)} (\tilde{y}_t - \hat{\mu}_i)(\tilde{y}_t - \hat{\mu}_i)', \tag{16}$$

where index sets $I(i)$ are given in (14).

Step III. *Estimation of the Markov matrix.* Given filtered data $\tilde{y}_t$ and regime estimates (15-16), select the Markov matrix P that minimizes Brier's score S, ie. the mean square error:

$$S = \frac{1}{T} \sum_t \left(\hat{P}\{J_t = 1 \mid \mathbf{Y}_t, \Theta\} - \delta(J_{t+l}, 1) \right)^2, \tag{17}$$

where the expansion regime probability $\hat{P}\{J_t = 1 \mid \mathbf{Y}_t, \Theta\}$ is computed recursively using (11-12), and the Kronecker's function $\delta(J_{t+l}, 1)$ is one if $J_{t+l} = 1$, otherwise zero.

Step II is very simple because the method ignores autocorrelation. This choice is supported by results in Layton (1996) and Lahiri and Wang (1994,1996), and is consistent with parsimony requirements (cf. Chatfield, 1996), when dealing with short time series, that additionally can be difficult to align. In the case of long and in phase correlated series, autoregressive terms could result in a better classification. They could be estimated ex post with known regimes in Step II, but the interpretation would become more complicated.

Lahiri and Wang (1994,1996) have emphasised that square errors around turning points are overcompensated for accuracy over expansions, leading to delayed signals. Other cost functions, eg. $C(e) = e^q, (q > 2)$, can be used in Step III if one wants a higher penalty for large errors occuring at turning points.

An interesting connection to Layton (1996) and Lahiri and Wang (1994,1996) is the following. Fixing their "quasi Bayesian" parameters produces the univariate and contemporaneous case of our method, but the inference on turning points is different. In the present method, decision rule (5) results in a *turning point signal* if the threshold level 1/2 is exceeded. This is an advantage over Lahiri and Wang (1994,1996), where the lack of feature extraction leads to a heuristically defined threshold (0.9), and Layton's (1996) rule: five propabilities in a row exceed 1/2.

We emphasize the importance of testing model forecasts outside the sample. Likelihood measures are not appropriate and conventional model selection criteria that would prevent overfitting, do not apply.

4. EMPIRICAL RESULTS

4.1 Data

As an application of the method outlined above, let's try to construct a model that signals the probability of a turning point in quarterly differences of log. Swedish industrial production (IP) in the next quarter (lead $l = 1$), according to the Swedish

National Accounts. For leading information we use two sources:

(i) The Swedish Business Tendency Survey (BTS): The balance between answers "higher" and "lower" expected production during the present quarter. This is published one quarter before IP which allows us to set lead equal to one.

(ii) The quarterly Stockholm Stock Market Index (SSMI).

Occasionally, BTS and SSMI don't correlate positively with industrial production, but jointly the three series work as a leading vector, with little risk of false turning point signals. The in-sample period, where the model is estimated, is 1971:Q1-1989:Q3 and post-sample is 1989:Q4-1997:Q2. The MBC programming has been done in Matlab[c]. A *turning point* between one regime and another is said to have occurred if the industrial production (Q4. diff. log.) has changed the sign and then has kept that sign for at least two quarters. All past turning points are defined in advance according to this rule. Figure 1 shows the IP time series and its regimes of recessions and expansions. Vertical lines indicate turning points.

4.2 Feature Extraction and Estimation

Both statistical and BTS data suffer from large errors, cf. Öller and Tallbom (1996). Hence there is a risk of false regime shift signals if the data were to be used as such. In Figure 2 the data is unfiltered. The regime is indicated as in Figure 1. The curve gives the probability of being in an expansion in the next quarter. This curve should preceed the regime indicator. We see in Figure 2 that an unfiltered indicator is not very reliable: two late signals and a varying lead time. Now apply exponential smoothing

$$\tilde{x}_t = \lambda x_t + (1-\lambda)\tilde{x}_{t-1},\ 0 < \lambda \leq 1, \tag{18}$$

to all series and a grid search $\lambda = .1, .2, \ldots, 1$. When the lead time was set to one, minimal Brier's scores were obtained for $\lambda = .2$ for IP, .7 for BTS, and .1 for SSMI. The large value of λ for BTS ensures that high amplitude signals get through without much delay, cf. Öller (1986). Markov probabilities were estimated using a

.05 grid for lambdas and .01 for finetuning the probability estimates. The estimates obtained were $\hat{p}_{11} = .94$ and $\hat{p}_{22} = .44$, revealing high asymmetry in persistence of regimes. In Figure 3, classification error probability bounds are calculated using (9). Note that the bounds become much narrower when the data are filtered, hinting of a reduction in the risk of making wrong inference. According to bounds (9), the maximum classification error probability is achieved for the prior values closest to $p = 1/2$. The Markov probability estimates imply that those values are achieved in recession regimes.

The method of Hamilton (1987) involves estimating model parameters by maximizing the likelihood function

$$L(\Theta \mid \mathbf{Y}_T) = \prod_{t=1}^{T} f(y_t \mid \Theta), \tag{19}$$

where $f(y_t \mid \Theta)$ is given by (13). Figure 4 shows that the resulting indicator would have been worthless if instead of (17) we wold have used (19).

4.3 Forecasting

The models were calibrated to lead $l = 1$. If getting early signals is of primary importance, an $l = 2$ calibration can be used, but then leads vary between zero and two, and a false turning point is signaled in 1996 (not shown here). Table 1 presents summary statistics of filtered and unfiltered MBC, the indicator using ML, and of a naive forecast.

The best MBC (filtered) had in-sample and out-of-sample Brier's scores .019 and .008, respectively. A standard naive competitor, the historical fraction of quarters for which the economy was in an expansion (here .680), had Brier's scores more than 10 times higher. Out-of-sample Brier's scores are better than in-sample, probably due to a sharp recession at the beginning of the 1990's which is advantageous to MBC models.

MBC leads to almost perfect inference, as seen from Figures 5-6. When the forecasting horizon is one quarter, MBC issues no false signals in the sense that the indicator would pass the one half probability line when there follows no change of regimes, or fail to pass the threshod when there follows a change. The only irregularity is 1972:Q1 where the lead is two quarters, instead of one.

Table 1. Summary Statistics for Sweden (lead $l = 1$).

Model	Smoothing Constant	Markov Probability	In-Sample Brier's S.	Out-of-S. Brier's S.
Filtered	.2, .7, .1	.94, .44	.019	.008
Unfilterd	-	.97, .53	.055	.016
ML+Filter	.1, .1, .1	.97, .93	.174	.168
Naive	-	-	.218	.242

4.4 The US Business-Cycle

Here the NBER-dated recessions are forecasted one quarter ahead ($l = 1$) by the quarterly per cent change in US GDP and the Composite Index of Leading Indicators (CLI) of the Department of Commerce. NBER-dates are reported in an appendix of Gordon (1997). The in-sample period is 1953:Q2 - 1973:Q2 and out-of-sample 1973:Q3 - 1993:Q2. Summary statistics are shown in Table 2. The best turning point forecasts were obtained using smoothing parameter values .4 for GDP and .5 for CLI. The grid was .3, .4, ... ,1. We tried .1, .2, ... ,1, finding an in-sample Brier's score minimum at (.2, .2). However, the low values of the smoothing constants lead to late signals in turning points, and much weaker out-of-sample behavior, alerting for overfitting. The Markov probability estimates were

$\hat{p}_{11} = .78$ and $\hat{p}_{22} = .02$ (grid as for Sweden) reflecting pathologically low persistence in recessions as compared to both Swedish figures and unfiltered MBC for the US, but out-of-sample behavior does not alert for overfitting. Note, however that the limit probabilities are .82 and .18, quite reasonable values.

The MBC-indicator is shown in Figure 7 and the corresponding decisions on regimes are illustrated in Figure 8. We see that the best MBC produces a turning point indicator that works in all turning points, except for the recession 1960:Q2 - 1960:Q4. The peaks are detected with leads varying between 0-4 quarters, the median lead being zero. The throughs are detected with leads between 0-3 quarters, the median lead being one. Again, MBC using filtered data produces the most reliable indicator. If no filtering would have been used, the median lead for trough detection would be zero. For the US data a lead does not include a publication delay, as in the Swedish case. Here, too, error bounds (9) become narrower for filtered data, especially in the recession regime. This corroborates the finding that the indicator is more reliable in detecting troughs than peaks.

Table 2. Summary Statistics for the US (lead $l = 1$).

Model	Smoothing Constant	Markov Probability	In-Sample Brier's S.	Out-of-S. Brier's S.	Median lead Peaks,Throughs:
Filtered	.4, .5	.78, .02	.064	.063	0, 1
Unfilterd	-	.95, .53	.071	.080	0, 0
Naive	-	-	.126	.136	-

5. DISCUSSION

We have introduced a new way to use HMM as a Markov-Bayesian classifier, where forecast accuracy, as measured by Brier's score, has been the minimization criterion. The proposed pattern recognition method provides a simple way to utilize leading information for forecasting regime shifts. In the empirical applications, leading information came from two expectational variables, the BTS and the SSMI, in the Swedish case, and from the CLI in the US case.

Past turning points were determined in advance. A key idea of this study is to utilize this easily available information. This and adopting a probability score made it possible to apply feature extraction which proved necessary for noisy data. Diebold and Rudebush (1991) found a substantial deterioration of forecasting performance in the preliminary figures. We had access only to revised data, but the feature extraction approach is well suited for even noisier preliminary figures.

The regime parameters are calibrated with the purpose of eliciting leading information. Another feature missing from earlier indicators based on HMM is a reliability measure. Here, well known classification error bounds are applied.

Our method does not lend itself to the task of determining past turning points. We also want to emphasize that it is not real Bayesian, but empirical Bayesian, since all parameters are estimated from data.

An almost perfect indicator for Swedish data is obtained for lead time one quarter. Due to publication lags this means contemporaneous in real time, but this is still a considerable gain. In the case of the US, MBC produced a reasonably working indicator with a one quarter lead. In both applications a feature extraction stage proved essential.

The method presented here was developed as a solution to a practical problem of forecasting business cycle turning points. However, we want to emphasize that it can be applied more generally as a dynamic classifier. Another typical application would be a noisy multivariate signal from a production process that occasionally goes into a state of malfunction.

References

Artis, M.J., Bladen-Hovell, R.C., Osborn, D.R., Smith, G.W. and Zhang, W. (1995), "Turning Point Prediction for the UK Using CSO Leading Indicators," *Oxford Economic Papers*, 47, 397-417.

Boldin, M.C. (1994), "Dating Turning Points in the Business Cycle," *Journal of Business*, 67, 97-131.

Chatfield, C. (1996),"Model Uncertainty and Forecast Accuracy," *Journal of Forecasting*, 15, 495-508.

Diebold, F.X., and Rudebusch, G. (1990), "A Nonparmetric Investigation of Duration Dependence in the American Business Cycle," *Journal of Political Economy*, 98, 596-616.

Diebold, F.X. and Rudebusch, G. (1991), "Forecasting Output With the Composite Leading Index: A Real Time Analysis," *Journal of the American Statistical Association*, 86, 603-610.

Funke, N. (1997), "Predicting Recessions: Some Evidence For Germany," *Weltwirtschaftliches Archiv*,133, 91-102.

Fukanaga, K. (1990), *Introduction to Statistical Pattern Recognition*, Academic Press.

Geisser, S. and Eddy, F.W. (1979),"A Predictive Approach to Model Selection," *Journal of the American Statistical Association*, 74, 153-168.

Gordon, S. (1997), "Stochastic Trends, Deterministic Trends, and Business Cycle Turning Points," *Journal of Applied Econometrics*, 12, 411-434.

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. (1994), *Time Series Analysis*, Princeton, N.J., Princeton University Press.

Hamilton, J.D. and Perez-Quiros, G. (1996a), "What Do Leading Indicators Lead?," *Journal of Business*, 69, 27-49.

Hamilton, J.D. and Lin, G. (1996b), "Stock Market Volatility and the Business Cycle," *Journal of Applied Econometrics*, 11, 573-593.

Lahiri, K. and Wang, J.G. (1994), "Predicting cyclical Turning points with Leading Index in a Markov Switching Model," *Journal of Forecasting*, 13, 245-263.

Lahiri, K. and Wang, J.G. (1996), "Interest Rate Spreads as Predictors of Business Cycles," *Handbook of Statistics*, 14, 297-315.

Layton, A.P. (1996), "Dating and Predicting Phase Changes in the U.S. Business Cycle," *International Journal of Forecasting*, 12, 417-428.

Leitch, G. and Tanner, J.E. (1995): "Professional Economic Forecasts: Are They Worth Their Costs?," *Journal of Forecasting*, 14, 143-157.

Lindgren, G., (1978): "Markov Regime Models for Mixed Distributions and Switching Regression," *Scandinavian Journal of Statistics*, 5, 81-89.

McCulloch, R.E. and Tsay, R.S. (1993), "Bayesian Inference and Prediction for Mean and Variance Shifts in Autoregressive Time Series," *Journal of the American Statistical Association*, 423, 968-978.

McCulloch, R.E. and Tsay, R.S. (1994), "Statistical Analysis of Time Series via Markov Switching Models," *Journal of Time Series Analysis*, 15, 523-539.

McNees, S. (1991).,"Forecasting Cyclical Turning Points (1991): The Record in the Past Three Recessions," in *Leading Economic Indicators, New Approaches and Forecasting Records* (eds.) Lahiri, K., and Moore, G. H., Cambridge University Press.

Morgenstern, O. (1965), *On the Accuracy of Economic Observations*, Princeton University Press.

Neftci, S.N. (1982), "Optimal Prediction of Cyclical Downturns," *Journal of Economic Dynamics and Control*, 4, 225-242.

Öller, L-E. (1986),"A note on Exponentially Smoothed Seasonal Differences," *Journal of Business and Economic Statistics*, 4, 485-489.

Öller, L-E. and Tallbom, C. (1996), "Smooth and Timely Business Cycle Indicators for Noisy Swedish Data," *International Journal of Forecasting*, 12, 389-402.

Shumway, R.H. (1982), "Discriminant Analysis for Time Series," *Handbook of Statistics*, 2, 1-46.

Stock, J.H. and Watson, M.W. (1989), "New Indexes of Coincident and Leading Economic Indicators," *NBER Macroeconomics Annual*, 351-394.

Stock, J.H. and Watson, M.W. (1993), *Business Cycles, Indicators, and Forecasting*, The University of Chicago Press.

Appendix: FIGURES

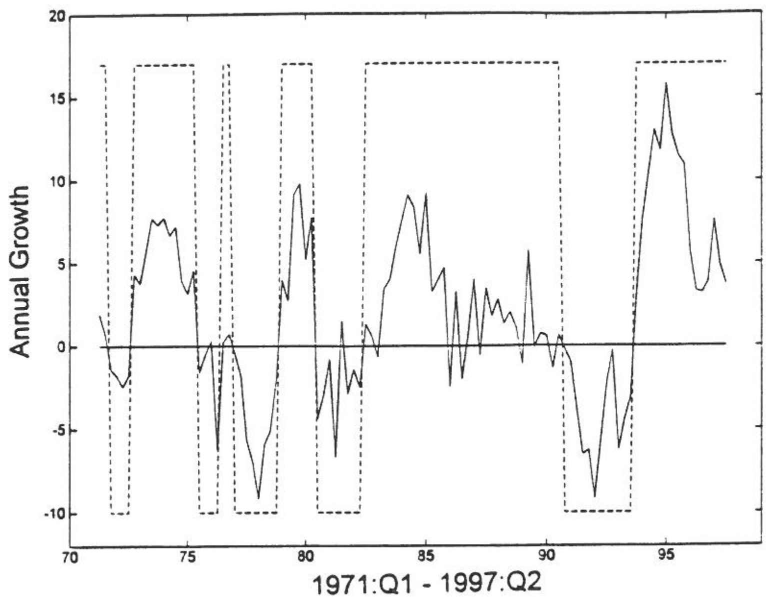

Figure 1. Swedish Industrial Production (Q4. diff. log) and Expansion/Recession Regimes.

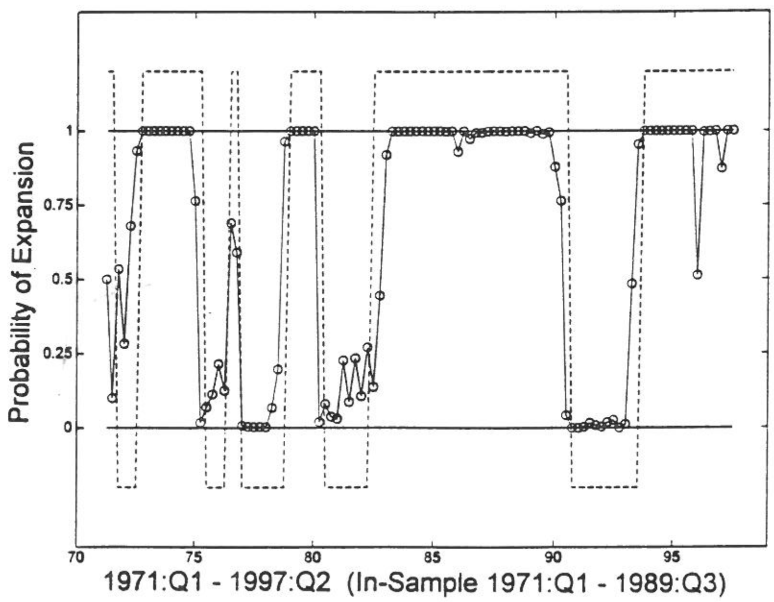

Figure 2. An Unfiltered Indicator for Sweden Based on Brier's Score (lead $l = 1$).

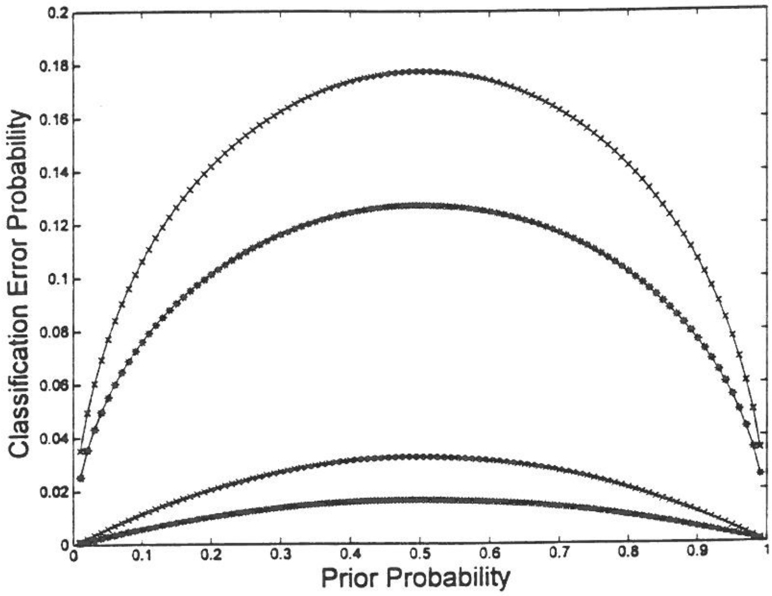

Figure 3. Bhattacharyya Classification Error Bounds for an Indicator Based on Filtered Data (***) and Unfiltered Data (+++).

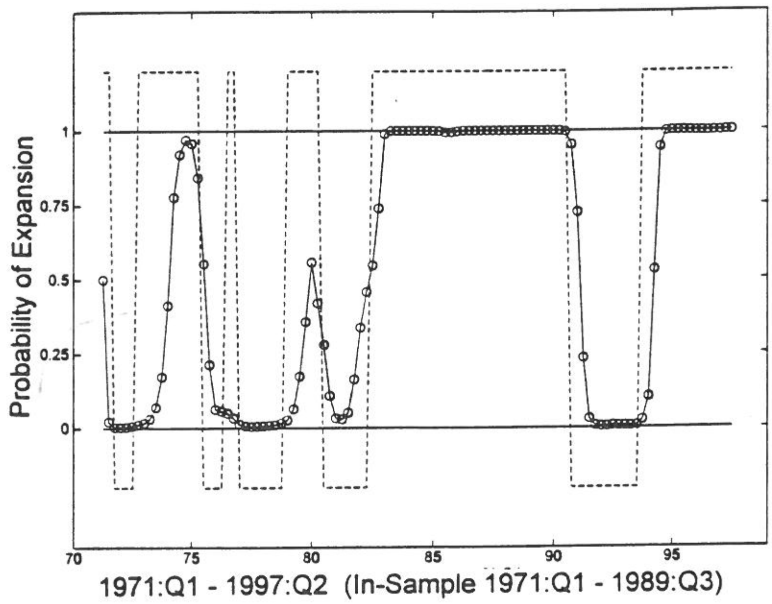

Figure 4. An Attempt at a Model Based on Maximum Likelihood and Feature Exraction.

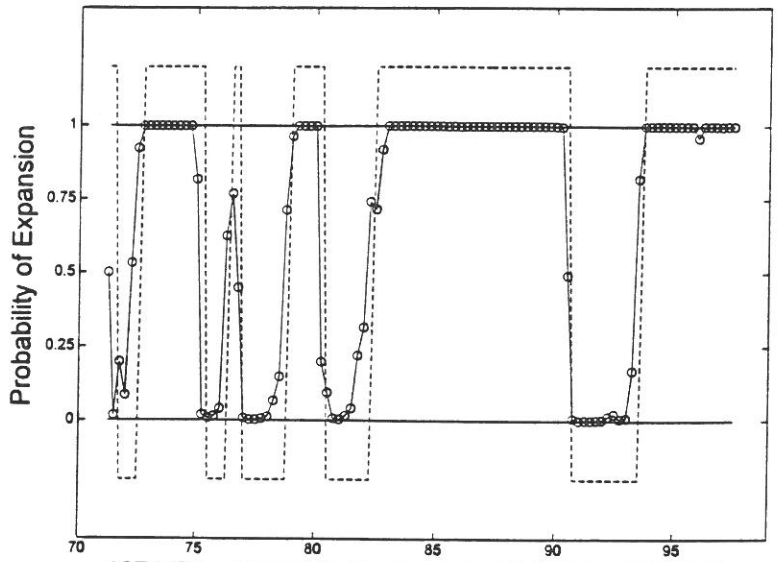

Figure 5. A Probability Indicator for Sweden, Based on Brier's Score and Feature Extraction (lead $l = 1$).

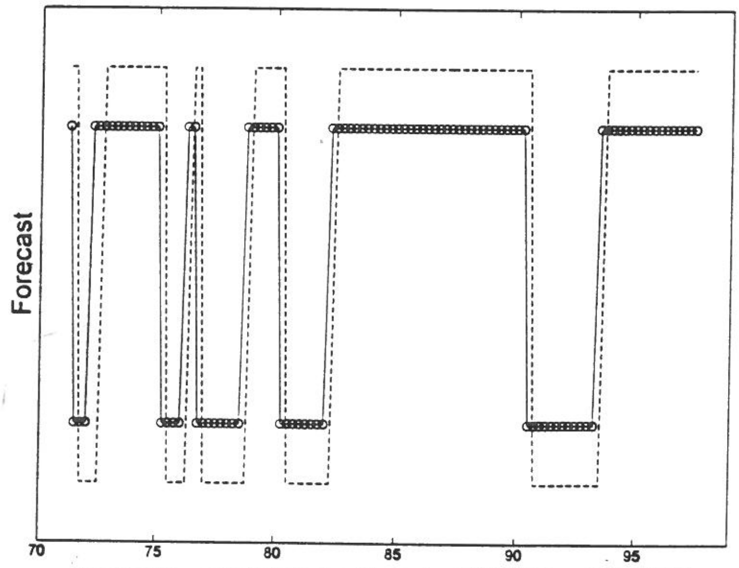

Figure 6. Inference Concerning Swedish Expansion/Recession Regimes Next Quarter, Using an Indicator Based on Brier's Score and Feature Extraction.

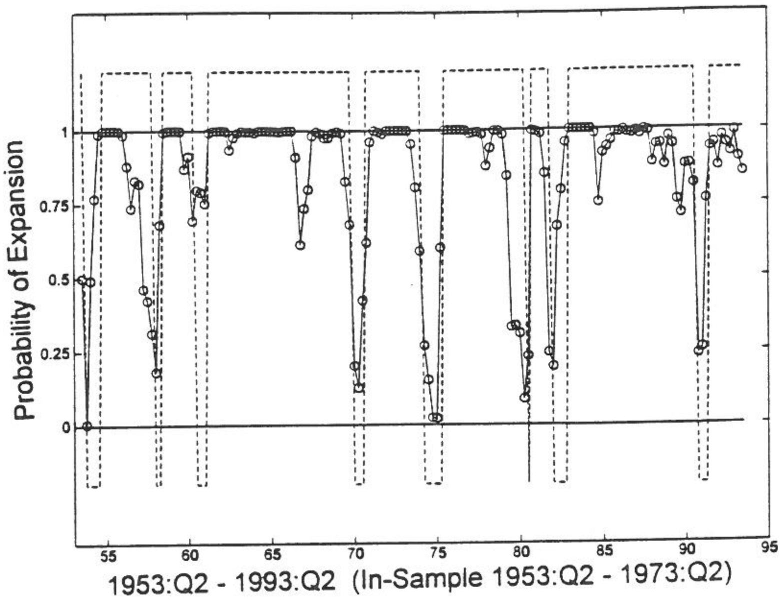

Figure 7. A Probability Indicator for the US, Based on Brier's Score and Feature Extraction (lead $l = 1$).

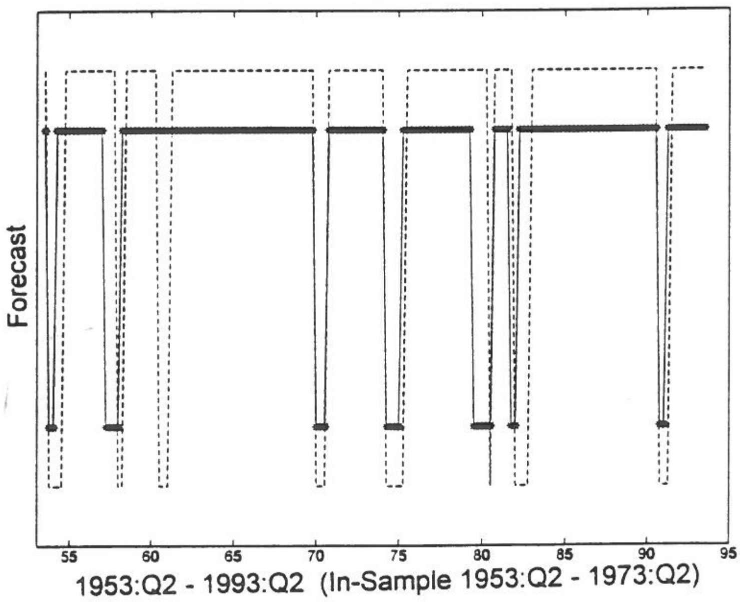

Figure 8. Inference Concerning US Expansion/Recession Regimes Next Quarter, Using an Indicator Based on Brier's Score and Feature Extraction.

Sammandrag

En dold Markov-modell använd som en dynamisk bayesiansk klassificerare och en tillämpning på vändpunktsprognoser

Att göra vändpunktsprognoser är både bergreppsmässigt och metodologiskt olikt det att man försöker prognosticera varje punkt i en tidsserie. Det är ofta mycket svårt och kräver stora mängder data om man vill bygga en modell som skall göra en prediktion på vartenda ett värde en tidsserie kommer att anta. Men är man bara intresserad av vändpunkter kan uppgiften göras enklare ifall man då fokuserar allt intresse enbart på dessa punkter. För ekonomiska agenter är det ofta viktigare att få korrekt information om utvecklingens riktning, än att erhålla så numeriskt noggranna prognoser som möjligt.

Den metod som utvecklas här bygger på *igenkänning av mönster* (pattern recognition). Vi utvecklar en probabilistisk klassificeringsmetod som är särskilt lämpad för att estimera sannolikheten för en konjunkturell vändpunkt. Tidsmässigt ledande information inhämtas från andra tidsserier. Ofta är det dock bara en viss del av den externa informationen som kan utnyttjas. I den vändpunktsindikator Konjunkturinstitutet nu använder utnyttjas enbart svängningar med låg frekvens eller hög amplitud, se Öller och Tallbom (1996). Även här användes samma typ av filtrering (extrahering av mönster), men filtreringsgraden bestämmes skilt för var serie, där objektfunktionen är prognosfelet. Samtidigt estimeras även Markov-sannolikheterna för att man skall stanna i samma regim (recession eller expansion). Markov-sannolikheternas väntevärden är emellertid konstanta och kan därför endast indirekt användas för att, via Bayes' formel (8) och uppdateringsformlerna (11) och (12), konstruera en indikator som varierar.

Det som skiljer vår metod från tidigare försök är att vi bestämmer de historiska vändpunkterna *a priori* enligt en enkel regel, de estimeras alltså inte. Därefter kalibreras modellen med hjälp av medelkvadratfelet i prognosen, kallat "Briers felpoäng" (17). För första gången ifråga om sannolikhetsindikatorer är vi istånd att ge felgränser (9) för prognossannolikheterna. Metoden testas på nationalräkenskapernas industriproduktion, där man utnyttjar två förväntningsvariabler, en serie från industribarometern och så aktieindex. En indikator erhålles som inte missar en enda vändpunkt varken under estimeringsperioden eller utanför denna. Inga felaktiga signaler förekommer heller, såsom lätt kunde ske vid den lilla svackan 1996.

Metoden testas även på USA:s industriproduktion där ledande information tas ur index för tolv ledande indikatorer. Resultatet är också i detta fall tillfredsställande.

Titles in the Working Paper Series

Number	Author	Title	Date
1	Warne, Anders and Anders Vredin	Current Account and Business Cycles: Stylized Facts for Sweden	12/89
2	Östblom, Göran	Change in Technical Structure of the Swedish Economy	12/89
3	Söderlind, Paul	Mamtax. A Dynamic CGE Model for Tax Reform Simulations	12/89
4	Kanis, Alfred and Aleksander Markowski	The Supply Side of the Econometric Model of the NIER	11/90
5	Berg, Lennart	The Financial Sector in the SNEPQ Model	2/91
6	Ågren, Anders and Bo Jonsson	Consumer Attitudes, Buying Intentions and Consumption Expenditures. An Analysis of the Swedish Household Survey Data	4/91
7	Berg, Lennart and Reinhold Bergström	A Quarterly Consumption Function for Sweden 1979-1989	10/91
8	Öller, Lars-Erik	Good Business Cycle Forecasts - A Must for Stabilization Policies	2/92
9	Jonsson, Bo and Anders Ågren	Forecasting Car Expenditures Using Household Survey Data	2/92
10	Löfgren, Karl-Gustaf, Bo Ranneby and Sara Sjöstedt	Forecasting the Business Cycle Not Using Minimum Autocorrelation Factors	2/92
11	Gerlach, Stefan	Current Quarter Forecasts of Swedish GNP Using Monthly Variables	2/92
12	Bergström, Reinhold	The Relationship Between Manufacturing Production and Different Business Survey Series in Sweden	2/92
13	Edlund, Per-Olov and Sune Karlsson	Forecasting the Swedish Unemployment Rate: VAR vs. Transfer Function Modelling	3/92
14	Rahiala, Markku and Timo Teräsvirta	Business Survey Data in Forecasting the Output of Swedish and Finnish Metal and Engineering Industries: A Kalman Filter Approach	3/92

15	Christoffersson, Anders, Roland Roberts and Ulla Eriksson	The Relationship Between Manufacturing and Various BTS Series in Sweden Illuminated by Frequency and Complex Demodulate Methods	4/92
16	Jonsson, Bo	Sample Based Proportions as Values on an Independent Variable in a Regression Model	10/92
17	Öller, Lars-Erik	Eliciting Turning Point Warnings From Business Surveys	11/92
18	Forster, Margaret M.	Volatility, Trading Mechanisms and International Cross-Listing	12/92
19	Jonsson, Bo	Prediction With a Linear Regression Model and Errors in a Regressor	12/92
20	Gorton, Gary and Richard Rosen	Corporate Control, Portfolio Choice, and the Decline of Banking	1/93
21	Gustafsson, Claes-Håkan and Åke Holmén	The Index of Industrial Production - A Formal Description of the Process Behind It	2/93
22	Karlsson, Tohmas	A General Equilibrium Analysis of the Swedish Tax Reforms 1989-1991	2/93
23	Jonsson, Bo	Forecasting Car Expenditures Using Household Survey Data - A Comparison of Different Predictors	2/93
24	Gennotte, Gerard and Hayne Leland	Low Margins, Derivative Securities, and Volatility	2/93
25	Boot, Arnoud W.A. and Stuart I. Greenbaum	Discretion in the Regulation of U.S. Banking	4/93
26	Spiegel, Matthew and Deane J. Seppi	Does Round-the-Clock Trading Result in Pareto Improvements?	4/93
27	Seppi, Deane J.	How Important are Block Trades in the Price Discovery Process?	4/93
28	Glosten, Lawrence R	Equilibrium in an Electronic Open Limit Order Book	4/93
29	Boot, Arnoud W.A., Stuart I. Greenbaum and Anjan V. Thakor	Reputation and Discretion in Financial Contracting	4/93
30a	Bergström, Reinhold	The Full Tricotomous Scale Compared with Net Balances in Qualitative Business Survey Data - Experiences from the Swedish Business Tendency Surveys	5/93

30b	Bergström, Reinhold	Quantitative Production Series Compared with Qualitative Business Survey Series for Five Sectors of the Swedish Manufacturing Industry	5/93
31	Lin, Chien-Fu Jeff and Timo Teräsvirta	Testing the Constancy of Regression Parameters Against Continuous Change	5/93
32	Markowski, Aleksander and Parameswar Nandakumar	A Long-Run Equilibrium Model for Sweden. The Theory Behind the Long-Run Solution to the Econometric Model KOSMOS	9/93
33	Markowski, Aleksander and Tony Persson	Capital Rental Cost and the Adjustment for the Effects of the Investment Fund System in the Econometric Model KOSMOS	9/93
34	Kanis, Alfred and Bharat Barot	On Determinants on Private Consumption in Sweden	10/93
35	Kääntä, Pekka and Christer Tallbom	Using Business Survey Data for Forecasting Swedish Quantitative Business Cycle Variables. A Kalman Filter Approach	11/93
36	Ohlsson, Henry and Anders Vredin	Political Cycles and Cyclical Policies. A new test approach using fiscal forecasts	11/93
37	Markowski, Alexander and Lars Ernsäter	The Supply Side in the Econometric Model Kosmos	1/94
38	Gustafsson, Claes-Håkan	On the Consistency of Data on Production, Deliveries, and Inventories in the Swedish Manufacturing Industry	8/94
39	Rahiala, Markku and Tapani Kovalainen	Modelling Wages Subject to Both Contracted Increments and Drift by Means of a Simultaneous-Equations Model With Non-Standard Error Structure	8/94
40	Öller, Lars-Erik and Christer Tallbom	Hybrid Indicators for the Swedish Economy Based on Noisy Statistical Data and the Business Tendency Survey	10/94
41	Östblom, Göran	A Converging Triangularization Algorithm and the Intertemporal Similarity of Production Structures	10/94
42a	Markowski, Aleksander	Labour Suply, Hours Worked and Unemployment in the Econometric Model Kosmos	10/94
42b	Markowski, Aleksander	Wage Rate Determination in the Econometric Model Kosmos	10/94 Rev.12/94

43	Ahlroth, Sofia, Anders Björklund and Anders Forslund	The Output of the Swedish Education Sector	12/94
44a	Markowski, Aleksander	Private Consumption Expenditure in the Econometric Model Kosmos	12/94
44b	Markowski, Aleksander	The Input-Output Core: Determination of Inventory Investment and Other Business Output in the Econometric Model Kosmos	12/94
45	Bergström, Reinhold	The Accuracy of the Swedish National Budget Forecasts 1955-92	3/95
46	Sjöö, Boo	Dynamic Adjustment and Long-Run Economic Stability	5/95
47a	Markowski, Aleksander	Determination of the Effective Exchange Rate in the Econometric Model Kosmos	7/95
47b	Markowski, Aleksander	Interest Rate Determination in the Econometric Model Kosmos	7/95
48	Barot, Bharat	Estimating the Effects of Wealth, Interest Rates and Unemployment on Private Consumption in Sweden	10/95
49	Lundvik, Petter	Generational Accounting in a Small Open Economy	1/96
50	Eriksson, Kimmo, Johan Karlander and Lars-Erik Öller	Hierarchical Assignments: Stability and Fairness	3/96
51	Url, Thomas	Internationalists, Regionalists, or Eurocentrists	4/96
52	Ruist, Erik	Temporal Aggregation of an Econometric Equation	10/96
53	Markowski, Aleksander	The Financial Block in the Econometric Model Kosmos	10/96
54	Göran Östblom	Emissions to the Air and the Allocation of GDP: Medium Term Projections for Sweden in Conflict with the Goals of CO_2, SO_2 and NO_x Emissions for Year 2000	11/96
55	Koskinen, Lasse, Aleksander Markowski, Parameswar Nandakumar	Three Seminar Papers on Output Gap	5/97
56	Oke, Timothy and Lars-Erik Öller	Testing for Short Memory in a Varma Process	5/97

57	Johansson, Anders and Karl-Markus Modén	Investment Plan Revisions and Share Price Volatility	12/97
58	Lyhagen, Johan	The Effect of Precautionary Saving on Consumption in Sweden	12/97