A CLASSIFYING PROCEDURE FOR SIGNALING TURNING POINTS

LASSE KOSKINEN[1] and LARS-ERIK ÖLLER[2]

[1] *The Central Pension Security Institute, PL 00065, Helsinki, Finland.* [2] *National Institute of Economic Research, P.O. Box 3116, S-103 62, Stockholm, Sweden, and Stockholm School of Economics.*

E-mail: *lasse.koskinen@etk.fi, lars-erik.oller@konj.se*

Abstract. A Hidden Markov Model (HMM) is used to classify an out of sample observation vector into either of two regimes. This leads to a procedure for making probability forecasts for changes of regimes in a time series, i.e. for turning points. Instead of maximizing a likelihood, the model is estimated with respect to known past regimes. This makes it possible to perform feature extraction and estimation for different forecasting horizons. The inference aspect is emphasized by including a penalty for a wrong decision in the cost function. The method is tested by forecasting turning points in the Swedish and US economies, using leading data. Clear and early turning point signals are obtained, contrasting favourable with earlier HMM studies. Some theoretical arguments for this are given.

Keywords. *Business Cycle, Feature Extraction, Hidden Markov Switching-Regime Model, Leading Indicator, Probability Forecast.*

1. INTRODUCTION

Forecasting turning points is both conceptually and methodologically different from making point forecasts of a time series. As recently pointed out by Keilis-Borok et al. (2000), recessions are rare, non-linear and complicated events. The chance of success in forecasting turning points is greatest if the method concentrates on this purpose only and uses all available information on turning points. In this paper, turning point prediction is interpreted solely as a classification problem. We develop a procedure that is particularly suitable for producing forecasts of turning points, where leading information is taken from other series. 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 some other concept has to be chosen, such as the probability of an occurrence of e.g. a turning point and risk bounds for false decision. Probability forecasts have been common in meteorology for several decades, but their appearance in economics is of more recent origin. Starting with a pioneering article by Neftci (1982), a number of probability methods for turning point forecasting has been suggested in the literature. Most of these are based either on Neftci's or Hamilton's (1989) business cycle models. Neftci's model uses sequential probabilities. Hamilton's probabilistic business-cycle model adopts the *Hidden Markov (Switching-Regime) Model* (HMM) from Lindgren (1978). Recently, HMM has been applied in constructing leading indicators, as suggested by Layton (1996), Lahiri and Wang (1994,1996), Hamilton and Perez-Quiros (1996) and Ivanova et al. (2000).

Our general approach is close to Artis et al. (1996), where it is explained how pattern recognition can be used for predicting turning points, applying Neftci's model. Here, instead of Neftci's method we use HMM for the same purpose. Combining known methods, we construct a simple classification procedure for producing probability forecasts for turning points. Our approach leads to flexible modeling. Estimation can be

adapted to the forecasting horizon and focused on turning points. This is because, instead of maximizing the likelihood, we minimize a linear combination of Brier's probability score and a classification error count estimate. The classification procedure consists of three stages: feature extraction, classification and evaluation. For general rules, see Fukanaga (1972). Feature extraction by a causal filter is used in order to facilitate classification. The present paper applies an *exponentially weighted moving average* (EWMA), passing low frequency and high amplitude data, as in Öller and Tallbom (1996), and temporally aligning turning point signals from the component series.

The paper is organized as follows. The next section provides an interpretation of HMM as a dynamic classifier and some theoretical results. The third section proposes a new classification procedure based on the tools of the preceding section. 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 briefly introducing static classification, subsequently extending the concepts to time series. The results and concepts concerning the static classification in Section 2.1 can be found e.g. in Fukanaga (1990). Section 2.2 suggests a new interpretation of the state estimation task of HMM as a dynamic Bayesian classifying problem and provides some new theoretical results concerning Markov dynamics in classification.

2.1 A Static Bayesian Classifier

Consider two classes $i \in \{1, 2\}$ (a generalization to an arbitrary number of classes is straight-forward) and a vector y of data to be allocated into either of these classes. Formally, there is a pair $Z = (Y, J)$, where $Y \in \mathbf{R}^n$ is a random vector and $J \in \{1, 2\}$ is a random variable that assigns class information to Y. Only Y is observed, 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. In order to simplify the notation we define a complement function i^c such that $i^c = 1$ if $i = 2$, otherwise $i^c = 2$.

Let the observations have a multivariate Gaussian distribution, $y \sim N(\mu_i, V_i)$, where μ_i is the mean and V_i the covariance matrix of Y in class i. Let π_i denote the prior probability $P\{J = i\}$, $i = 1, 2$. The posterior probability for an observation y belonging to class 1 or 2 follows from *Bayes' theorem*:

$$P\{J = i \mid Y = y\} = \frac{\pi_i \times f(y \mid J = i)}{\pi_1 \times f(y \mid J = i) + \pi_2 \times f(y \mid J = i^c)}, \tag{1}$$

where $f(y \mid J = i)$ is the normal density of y, given class i. The *Gauss-Bayesian classifier* allocates y to class i, $g(y) = i$ if and only if

$$P\{J = i \mid Y = y\} > \frac{1}{2}. \tag{2}$$

We disregard from equality because it only holds with probability zero. Probabilities (1) express the uncertainty: the closer the posterior probability is to one or to zero, the less uncertainty there is in a decision. The central concept in judging the reliability of the classifier is the probability that a sample is assigned to the wrong class, that is, the *Probability of Error* (Overall Bayes Risk) $R\{Y\}$:

$$R\{Y\} = P\{g(Y) \neq J\}. \tag{3}$$

When Y is a random vector, the calculation of (3) is usually very complicated, but upper and lower bounds are given by

$$\frac{1}{2}\left(1 - \sqrt{1 - 4 \times P\{J = i \mid Y = y\} \times (1 - P\{J = i \mid Y = y\}) \, e^{-2B^2}}\right)$$

$$\leq R\{Y\} \leq e^{-B^2} \times \sqrt{P\{J = i \mid Y = y\} \times (1 - P\{J = i \mid Y = y\})}, \tag{4}$$

where the B^2 is the *Bhattacharya distance* between classes 1 and 2:

$$B^2 = \frac{1}{8}(y - \mu_1)'V^{-1}(y - \mu_2) + \frac{1}{2}log\left(\frac{|V|}{\sqrt{|V_1| \times |V_2|}}\right), \tag{5}$$

and $V = \left(V_1 + V_2\right)/2$, see Fukanaga (1972), Ch. 9.

3

2.2 A Dynamic Bayesian Classifier

The classifier defined in (1) and (2) could be called static. However, our aim is to construct a dynamic probability classifier that also describes the dependence between observations when dealing with time series data, $Z_t = (Y_t, J_t)$. The normality assumption of classes now takes the form:

$$Y_t \sim N(\mu_{J_t}, V_{J_t}). \tag{6}$$

In order to model time dependence in a mathematically tractable way, we postulate that J_t can be described as a first-order 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 (6). A Markov chain generates switching between classes. When in class i, the process is said to be working in *regime i*. Denote the *Markov Probability* that regime i will be followed by regime j by

$$p_{ij} = P\{J_t = j \mid J_{t-1} = i\}. \tag{7}$$

The model is now defined by regime distributions (6) and by the matrix $\mathbf{P}$ consisting of probabilities (7). The complete parameter set is $\Theta = \{\mu_1, \mu_2, V_1, V_2, p_{11}, p_{22}\}$. Note that the introduced model can be interpreted as a special case (without autoregressive terms) of HMM, where Y_t is an observed time series that depends on an hidden Markov chain J_t.

At $t-1$ we denote the posterior probabilties by $p_i(t-1) = P\{J_{t-1} = i \mid Y_{t-1} = y_{t-1}\}$. The prior probabilities for regimes at t can be expressed by Markov probabilities and previous posterior probabilities as

$$\pi_i(t) = p_{1i} \times p_1(t-1) + p_{2i} \times p_2(t-1) \tag{8}$$

and the posterior probabilities for given y_t at t are

$$p_i(t) = \frac{\pi_i(t) \times f(y_t \mid J_t = i)}{\sum_{j=1}^{2} \pi_j(t) \times f(y_t \mid J_t = j)} \tag{9}$$

4

where $f(y_t \mid J_t = i)$ is the normal density (6) of y_t, given $J_t = i$. Assumptions (6) and (7), formula (9) and decision rule (2) ($g(y_t) = i$ if and only if $P\{J_t = i \mid Y_t = y_t\} > \frac{1}{2}$) define a dynamic variant of the Gauss-Bayesian classifier that we call a *Markov-Bayesian Classifier* (MBC). According to (9), decision rule (2) can equivalently be expressed as

$$\pi_i(t) \times f(y_t \mid J_t = i) \ > \ (1 - \pi_i(t)) \times f(y_t \mid J_t = i^c) \tag{10}$$

or in likelihood ratios and odds:

$$-log\Big(\frac{f(y_t \mid J_t = i)}{f(y \mid J_t = i^c)}\Big) \ < \ log\Big(\frac{\pi_i(t)}{1 - \pi_i(t)}\Big). \tag{11}$$

This rule is used in Hamilton (1989) and it has the smallest probability of error (3) among all classifiers when (6) and (7) hold and $\pi_i(t)$, $(i = 1, 2)$ are known (see Fukanaga (1972), Ch. 3). Note that (11) is the formulation based on information theory suggested in Birchenhall et al. (1999).

The conditional probability of error of MBC depends heavily on the prior probability $\pi_i(t)$, as the following lemma shows.

Lemma 1. Denote the classifying function

$$T(y_t) = (y_t - \mu_i)'V_i^{-1}(y_t - \mu_i) - (y_t - \mu_{i^c})'V_{i^c}^{-1}(y_t - \mu_{i^c}) - log\Big(\frac{|V_{i^c}|}{|V_i|}\Big), \tag{12}$$

and the corresponding inference set

$$D_t = \Big\{y_t \mid T(y_t) > 2 \times log\Big(\frac{\pi_i(t)}{1 - \pi_i(t)}\Big)\Big\}. \tag{13}$$

Then the conditional probability of error of MBC at t is

$$R\{Y_t \mid J_t = i\} = \int_{D_t} f(y_t \mid J_t = i)dy. \tag{14}$$

Proof. MBC allocates y_t to regime i if and only if (11) holds. Since the regimes of MBC are Gaussian, according to (12), inequality (11) is equivalent to

$$T(y_t) < 2 \times log\Big(\frac{\pi_i(t)}{1 - \pi_i(t)}\Big). \tag{15}$$

5

Now, if $J_t = i$, we can conclude from formula (15) that a classification error occurs if and only if

$$T(y_t) > 2 \times log\left(\frac{\pi_i(t)}{1 - \pi_i(t)}\right). \tag{16}$$

Hence, the conditional probability of error is

$$R\{Y_t \mid J_t = i\} = \int_{D_t} f(y_t \mid J_t = i)dy, \tag{17}$$

which completes the proof.

When the parameter set Θ of MBC is known, well known estimation algorithms for the state estimation of HMM can be used for classification. Hamilton (1989) presents a least squares algorithm in a recursive form, utilizing Bayes' theorem and the Markov property as follows.

The posterior probability estimates at $t - 1$, $\hat{P}\{J_{t-1} = i \mid Y_{t-1} = y_{t-1}\}$, and the prior probability estimates (Markov predictions) at t, $\hat{P}\{J_t = i \mid Y_{t-1} = y_{t-1}\}$, are collected into a pair of 2×1 vectors. Then estimates for the posterior probabilities (regime probabilities) of future observations $\{y_t, y_{t+1}, y_{t+2}, \ldots\}$ can be found by iterating the following pair of equations:

$$\begin{pmatrix} \hat{P}\{J_t = 1 \mid Y_t = y_t\} \\ \hat{P}\{J_t = 2 \mid Y_t = y_t\} \end{pmatrix} = \frac{\begin{pmatrix} \hat{P}\{J_t = 1 \mid Y_{t-1} = y_{t-1}\} \times f(y_t \mid J_t = 1) \\ \hat{P}\{J_t = 2 \mid Y_{t-1} = y_{t-1}\} \times f(y_t \mid J_t = 2) \end{pmatrix}}{f(y_t \mid y_{t-1})}, \tag{18}$$

and

$$\begin{pmatrix} \hat{P}\{J_t = 1 \mid Y_{t-1} = y_{t-1}\} \\ \hat{P}\{J_t = 2 \mid Y_{t-1} = y_{t-1}\} \end{pmatrix} = \begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix}' \begin{pmatrix} \hat{P}\{J_{t-1} = 1 \mid Y_{t-1} = y_{t-1}\} \\ \hat{P}\{J_{t-1} = 2 \mid Y_{t-1} = y_{t-1}\} \end{pmatrix}, \tag{19}$$

where $f(y_t|y_{t-1})$ is

$$f(y_t \mid y_{t-1}) = \hat{P}\{J_t = 1 \mid Y_{t-1} = y_{t-1}\} \times f(y_t \mid J_t = 1)$$

$$+\hat{P}\{J_t = 2 \mid Y_{t-1} = y_{t-1}\}) \times f(y_t \mid J_t = 2). \tag{20}$$

In this paper we set the neutral starting probability of $1/2$ for the recursion. We return to the problem of estimating the model parameters in the next section.

As can be seen from (8) and (14), a direct implication of Lemma 1 is the following theorem.

Theorem 1. Assume that $\hat{P}\{J_{t-1} = i \mid Y_{t-1} = y_{t-1}\} = 1$. Then the following limits hold for the conditional probability of error of MBC at t:

$$\lim_{p_{ii} \longrightarrow 1} R\{Y_t \mid J_t = i\} = 0, \quad \lim_{p_{ii} \longrightarrow 1} R\{Y_t \mid J_t = i^c\} = 1, \tag{21}$$

$$\lim_{p_{ii} \longrightarrow 0} R\{Y_t \mid J_t = i\} = 1, \quad \lim_{p_{ii} \longrightarrow 0} R\{Y_t \mid J_t = i^c\} = 0. \tag{22}$$

Another implication of Lemma 1 concerns the structure of the model. Consider the effect of the Markov probability parameters in the estimation of regimes by MBC at a regime-switch and within a regime.

Theorem 2. Assume that $J_{t-1} = i$ and denote $\hat{p}_i(t-1) = \hat{P}\{J_{t-1} = i \mid Y_{t-1} = y_{t-1}\}$.

I) A regime-switch occurs at t.

a) If $p_{11} + p_{22} > 1$, then the smaller is the absolute estimation error $|\hat{p}_i(t-1) - 1|$, the larger is the conditional probability of error $R\{Y_t \mid J_t = i^c\}$ of MBC at t.

b) If $p_{11} + p_{22} < 1$, then the smaller is the absolute estimation error $|\hat{p}_i(t-1) - 1|$, the smaller is the conditional probability of error $R\{Y_t \mid J_t = i^c\}$ of MBC at t.

II) A regime-switch does not occur at t.

a) If $p_{11} + p_{22} > 1$, then the smaller is the absolute estimation error $|\hat{p}_i(t-1) - 1|$, the smaller is the conditional probability of error $R\{Y_t \mid J_t = i\}$ of MBC at t.

b) If $p_{11} + p_{22} < 1$, then the smaller is the absolute estimation error $|\hat{p}_i(t-1) - 1|$, the larger is the conditional probability of error $R\{Y_t \mid J_t = i\}$ of MBC at t.

Proof. We shall prove only case I a). Analogous proofs can be given for the remaining cases. Lemma 1 implies that if the odds, ie. the ratio of prior probabilities ($\pi_i(t) = p_{1i} \times \hat{p}_1(t-1) + p_{2i} \times \hat{p}_2(t-1)$)

$$\frac{\pi_i(t)}{1 - \pi_i(t)} \tag{23}$$

increases then so does the conditional probability of error $R\{Y_t \mid J_t = i^c\}$ of the complementary case. Hence, it is sufficient to show that (23) is an increasing function in $\hat{p}_1(t-1)$. This can be done by studying the derivative of (8):

$$D_{\hat{p}_1(t-1)}\left(\frac{\pi_i(t)}{1 - \pi_i(t)}\right) > 0 \tag{24}$$

if and only if $p_{11} + p_{22} > 1$; Case I a) is proved.

Remark. If $p_{11} + p_{22} = 1$, the previous estimate $\hat{P}\{J_{t-1} = 1 \mid Y_{t-1} = y_{t-1}\}$ has no effect on $\hat{P}\{J_t = 1 \mid Y_t = y_t\}$ and hence MBC is static.

The theorems indicate that MBC may behave paradoxically at the crucial point of a regime switch. Although the classifier works perfectly at $t-1$, it will almost certainly miss a regime switch ($J_t \neq J_{t-1} = i$), if p_{ii} approaches unity. Moreover, a small error at $t-1$ results in a large error at t if the sum of the probabilities of remaining in a regime ($p_{11} + p_{22}$) exceeds unity.

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, a Markov chain provides a reasonable description of the traditional NBER business cycle dates (cf. eg. Hamilton (1989), Diebold and Rudebush (1990)). Secondly, the use of several leading series is essential in turning point forecasting (cf. eg. Keilis-Borok et al. (2000)). The classification methodology supports the use of a vector series.

Algorithm (18) and (19) is used in model estimation with a cost function emphasizing inference rule (2). The data are first filtered in order to facilitate classification.

When estimating the parameters of a HMM, usually a maximum likelihood (ML) approach is chosen. We adopt a predictive approach to model specification, basing parameter estimation and model selection on a cost function that minimizes the risk of wrong inference concerning turning points. This provides a possibility for feature extraction and for calibrating the model to the forecasting horizon.

We propose a three stage data-driven procedure, where stages are iterated with different feature extraction filters and Markov matrices, until an optimal model, under a given cost function, is found. We assume that past regimes, *recessions* or *expansions*, are known. As a result, regime parameters are easily estimated and an appropriate cost function can be applied. The computational complexity of the method depends on the class of filters used (if any) and the optimization algorithm. However, computations are easy because no autoregressive terms exist and necessary moments are estimated from known regimes.

Let J_t denote the regime series, as shown by the dotted lines in Figures 1-3 and 5. Our task is to select a model that predicts J_{t+l}, $l > 0$ at t (with lead l), applying a filter and MBC on an observed series y_t, containing leading information. The observations are divided into an estimation period and a test period. The data from the latter period are not allowed to have any influence on estimation. Let the estimation sample be $\{y_0, y_1, \ldots, y_{T+l-1}\}$. Then, estimation is based on the errors e_t of regime probability estimates (18-19)

$$e_t = \hat{P}\{J_{t+l} = i \mid y_t\} - \delta(J_{t+l}, i), \tag{25}$$

where the Kronecker function $\delta(J_{t+l}, i)$ is one if $J_{t+l} = i$, and otherwise zero. Since both inference and its uncertainty are essential in classification, we shall measure the quality of the inference using the *Error Count Estimate* of the probability of error (3) in classification (the number of elements in a set is denoted by $\#$):

$$ECE = \frac{1}{T}\#\{\ y_t \mid g(y_t) \neq J_{t+l}\}, \tag{26}$$

9

and the uncertainty of inference using *Brier's Probability Score* ie. the mean square error:

$$MSE = \frac{1}{T} \sum_t e_t^2. \tag{27}$$

In fact, (26) is the more important criterion when evaluating turning point forecasts, but the rareness of recessions necessities (27) in model building. The model estimation procedure consists of a loop of three steps (within the estimation period):

Step I. *Feature extraction.* Apply a causal filter F on $\{y_0, y_1, \ldots, y_T\}$. We denote $\tilde{y}_t = F(y_t)$.

Step II. *Estimation of regime parameters.* Divide the dates $I = \{0, \ldots, T\}$ into two sets:

$$I(i) = \{t \in I \mid J_{t+l} = i\}, \quad i = 1, 2. \tag{28}$$

Then estimate means and covariance matrices by

$$\hat{\mu}_i = \frac{1}{\#\{I(i)\}} \sum_{t \in I(i)} \tilde{y}_t \tag{29}$$

$$\hat{V}_i = \frac{1}{\#\{I(i)\}} \sum_{t \in I(i)} (\tilde{y}_t - \hat{\mu}_i)(\tilde{y}_t - \hat{\mu}_i)'. \tag{30}$$

Step III. *Estimation of the Markov matrix.* Given filtered data for the sample period $\tilde{y}_t$ and regime estimates (29) and (30), compute the Markov probability estimates using (18) and (19). Then select the Markov matrix $\mathbf{P}$ that minimizes the cost function

$$S = \omega \times MSE + (1 - \omega) \times ECE \tag{31}$$

where $\omega \in (0, 1)$. The value of ω should be chosen according to the user's utility function and the data. In both case studies in Section 4 we set $\omega = 2/3$.

Steps I-III are repeated for grids of filter parameters and of the Markov probabilities. Estimates that minimize (31) are chosen as the preferred model. This model is then further tested using the data that were saved for this purpose.

We emphasize the importance of testing a model outside the estimation period. Our procedure is not based on the ML principle and hence conventional model selection criteria, which would prevent overfitting, do not apply. However, in our examples the combined cost function (31) protected against overfitting. Also, note that Swanson and White (1997) tested several econometric models on nine macroeconomic time series and found that *the estimation period Schwarz Information Criterion was not superior to true out-of-sample forecast measures for selecting models*. They, too, emphasize the importance of choosing an appropriate cost function.

A causal filter has two effects on a time series y_t. Firstly, it attenuates high frequency noise, reducing the risk of false alarms. Secondly, it *shifts the phase* of the series. The component series may have different leads with respect to the regimes. Then, a filter can also be used to put the component series into the same phase, maximising the strenght of the signal. An effect of cost function (31) is that a proper filter is chosen.

When a series with a correct lead is filtered there is trade-off effect. By reducing the variance, the error probability within regimes decreases, but filtering also shortens the distance between observations at turning points and, hence, it increases the error probability there. The Markov dynamics of MBC can be expected to change dramatically due to filtering, because by Theorems 1 and 2, a low probability to stay in a regime improves turning point detection, thus compensating for the increased error probability due to the shortened distance between observations.

3.2. Relations to previous HMM studies

Estimating a HMM by ML, Lahiri and Wang (1994,1996) emphasize that large square errors (27) around turning points are overcompensated for high accuracy during long expansion periods. In other words, the ML criterion favours high probabilities of remaining in a regime, because regime shifts are rare events. Theorems 1 and 2 make us expect that this results in late turning point signals. This tendency can be seen eg. in Hamilton and Perez-Quiros (1996). The error count estimate (28) imposes an extra penalty for errors larger than 1/2 (wrong inference), correcting for the overcompensa-

11

tion.

Hamilton (1989) and Hamilton and Perez-Quiros (1996) have proposed the use of (19) to forecast regimes. However, the transition matrix, (e.g. Filardo (1994) allows it to vary), is constant and it works as a linear operator. For our method, it is crucial that one estimates the model for horizon l, and *elicits information* from y_t on J_{t+l}. For leading indicators, the same observation is made in Estrella and Mishkin (1998), and in Cox (1961), in the context of a forecast generated by EWMA, when in fact the data generator is not the corresponding integrated moving average process.

An interesting connection to Layton (1996) and Lahiri and Wang (1994,1996) is the following. Fixing their "quasi Bayesian" parameters produces the univariate, non-smoothed and contemporaneous case of our method, but the inference on turning points is different from (2). We will return to this in Section 4.2. In the present method, decision rule (2) results in a *turning point signal* when the probability estimate (18) exceeds 1/2. This is an advantage over Lahiri and Wang, and Layton (1996), where the lack of feature extraction leads to a heuristically defined threshold (0.9) in the former, and an *ad hoc* rule: "five probabilities in a row exceed 1/2", for monthly data, in the latter. Hamilton and Perez-Quiros (1996) ignore the inference aspect when evaluating turning point forecasts.

The method does not exploit autocorrelation. In case of strongly autocorrelated time series, this is an obvious shortcoming. Our choice is supported by results in Layton and Lahiri and Wang, and is consistent with parsimony requirements (cf. Chatfield, 1996), when dealing with short time series, that additionally can be difficult to align temporally. Secondly, as stated in Ivanova et al. (2000), the effects of autoregressive parameters will largely be captured by the probabilities of remaining in the current state.

Others that have used HMM endeavour to achieve four goals simultaneously: description of the time series, definition of turning points, and making point and probability forecasts. To summarize, the MBC procedure has only one aim: to produce accurate probability forecasts of turning points. MBC can be justified in five ways:

(i) The information on past turning points is used in estimation.

(ii) The model can be optimized with respect to any user-defined turning point.

(iii) A clear decision rule is explicitly used in estimation.

(iv) Feature extraction may improve resolution essentially.

(v) MBC is simple and computationally easy.

4. EMPIRICAL RESULTS

We now apply MBC to construct a leading probability indicator for Sweden and the USA. For Sweden we scanned for potential variables among those that were quick to appear and reflect expectations, selecting the Business Tendency Survey and the Stockholm Stock Exchange Index series. The reference (coincident) series is Industrial Production (in the National Accounts). In the case of the USA, we use GDP and the Composite Index of Leading Indicators (CLI) of the Department of Commerce. In both applications we divided the data into an estimation and a testing period. In the latter the data were not allowed to influence estimation.

For feature extraction we apply EWMA. Smoothing parameters and Markov probabilities are determined in two stages of grid search. In the first stage, the EWMA parameter $\lambda = .1, .2, \ldots, 1$ and $p_{ii} = .05, .1, .15, .., 1$, $i = 1, 2$. In the second stage, Markov probabilities are finetuned by a grid with step length .01. The estimation was done in Matlab$^{\copyright}$.

4.1 Sweden's Indicator

We construct a model that signals the probability of a turning point in Swedish Industrial Production in the next quarter (a "lead one" indicator). The publication of statistics occurs with a lag of one quarter. Hence, in real time, lead and lag cancel. For leading information we use two sources:

(i) The Swedish *Business Tendency Survey* (BTS): The balance between answers "higher" and "lower" production during the present quarter.

13

(ii) The quarterly differences ($\triangle_4$) of the logarithm of *the Stockholm Stock Exchange Index* (SSEI).

The observations 1971 Q1 to 1989 Q3 were used for estimation and 1989 Q4 to 1998 Q2 for testing the model. The serious recession in Sweden, which can be seen in Figure 1, at the beginning of the 1990s, provides a tough test of model accuracy and should be included in the test period. On the other hand, the series are too short for starting the testing before 1989.

We used a seasonally unadjusted reference series IP. It is nonstationary and hence needs differencing for the moments to exist. By choosing quarterly differences we take care of both seasonality and non-stationarity. Differencing also robustifies against structural breaks, see Clements and Hendry 1999, Ch. 5.

New data on BTS and SSEI are published one quarter before IP which allows us to set lead equal to one. The latest available observation vector at quarter t has the form $y_t = (\triangle_4 \ log \ IP_{t-1}, \ BTS_t, \ \triangle_4 \ log \ SSEI_t)$ and a forecast with lead time one is approximately contemporaneous in real time. An example would be the following. In the middle of April, observations of BTS and SSEI for the first quarter become available, but the latest observation on IP is from the fourth quarter of the previous year.

A turning point is said to have occurred when the reference series is in the expansion regime, then becomes negative and stays negative during the next quarter. The first quarter with the changed sign is called a *turning point* and vice versa for a contraction. This is *Oku's two sign rule*. An example is the following time series, $\{1, 2, -\mathbf{1}, -2, \mathbf{3}, 4\}$, where bold figures indicate turning points. Figure 1 shows the IP series. Vertical parts in the dotted line indicate turning points, horizontal regimes.

Occasionally, BTS and SSEI do not correlate positively with industrial production, but the joint vector y_t carries leading information with little risk of a false turning point signal. When the lead time was set to one in Table I, minimum (31) scores were obtained for $\lambda = .2, .7$ and $.1$ for IP, BTS, and SSEI, respectively. The large value of λ for BTS ensures that high amplitude signals get through without much delay. The effects of different values of λ are given in Öller (1986). The extremely low $\lambda = .1$ for

SSEI means that only a long lasting and strong downswing in stock prices increases the recession probability. The strong smoothing can be afforded because of long lead in stock prices; the loss function (31) chooses a phase shift that aligns this variable with the rest. This supports good resolution.

The MBC, calculated as described in Section 3, was compared to several alternative models in Table I. A standard naive competitor N, is the historical fraction (.68) of quarters for which the economy was in expansion. MBC-UF is the unfiltered MBC. The next two models do not apply any loss criterion; they are entirely based on the characteristics of past (and known) regimes. HMM-STAT is the static (formula (1)) HMM whose transition probabilities are the fraction of quarters in each regime, $\hat{p}_{11} = .68, \hat{p}_{22} = .32$ (see the remark to Theorem 2). HMM-DYN is like the previous model, but it is dynamic; the Markov probabilities are estimated from the observed regimes.

The ranking of the models is obvious: the naive model N is the worst of all and MBC is the best. This holds both within the sample and outside it, as well as for both MSE and ECE. MBC also has the lowest Bhattacharya upper bounds (4). The comparison between MBC and HMM-DYN is shown in Figure 4. Figure 2 shows MBC probability forecasts. The regimes are indicated as in Figure 1. The line with small circles shows the probability of being in an expansion in the next quarter. The corresponding inference on regimes is illustrated in Figure 3. Note that MBC leads to no wrong inference, neither within the sample nor in the test period! We have also experimented with a lead two indicator. This is not as accurate, but used in tandem with the lead one indicator it has proved useful. Here we used a forward looking BTS series and SSEI, while IP got a zero weight, because it simply contained no information on its own value half a year ahead. Also, the long lead of SSEI was corroborated.

4.2 The US Indicator

Here the NBER-dated US recessions are forecasted by the change in US GDP and the *Composite Index of Leading Indicators* (CLI) of the Department of Commerce. NBER-dates and the GDP series are reported in an appendix of Gordon (1997). As in

Hamilton and Perez-Quiros (1996), where the first difference ($\triangle$) of the logarithm of GNP and preliminary figures on CLI were used with the HMM, the estimation period is 1953:Q2 - 1973:Q2 and the test period is 1973:Q3 - 1993:Q2. We found that the results are not very sensitive to the precise date at which the sample period ends. The publication of CLI and GDP statistics occurs with a lag of one quarter. Hence the observation series is of the form $y_t = (\triangle \ log \ GDP_{t-1}, \ CLI_{t-1})$.

We started by using only CLI to forecast the NBER turning points. Table II reports the results of a comparison between the one-dimensional MBC and HMM with no autoregressive terms applying ML. This was the HMM used in Lahiri and Wang (1994, 1996) and in Layton (1996). For all lead times, MBC is an improvement on HMM, according to (31). Note that, for $l = 0$, MBC becomes almost static ($\hat{p}_{11} + \hat{p}_{22} = 1.01$), see the remark to Theorem 2.

In Table III, the bivariate MBC (GDP and CLI) was compared to the same models as in the previous section. The table presents summary statistics of each model, calibrated to lead $l = 0, 1, 2$, (ie. -1, 0, 1, in real time). Again, MBC was clearly the best. It produces an indicator that works in all turning points, except when entering the recession 1960:Q2 - 1960:Q4, for $l = 1$. This is an unusual achievement. The recession in the beginning of 90s has been especially difficult to forecast, see Fintzen and Stekler (1999). In Stock and Watson (1993), the 1990 recession was missed and Figure 1 in Hamilton and Perez-Quiros (1996) and Figures 1-2 in Birchenhall et al. (1999) show that either the indicators are late or they miss that recession completely. In Figure 5 a clear coincident warning is given. The recorded range of leads was the narrowest for MBC, as were the error the bounds (4) (not shown here). The MBC-indicator with lead one is shown in Figure 5. The scores of MBC-UF are no better than those of HMM-STAT and HMM-DYN, and definitely worse than for MBC. This again shows the importance of smoothing.

In Table III the smoothing constant λ for GDP falls sharply as the lead increases. At the same time turning point detection capability is increased by lowering $\hat{p}_{11} + \hat{p}_{22}$, see Theorem 1. When the lead is 1 and 2 the sum is smaller than unity. For the

16

unfiltered MBC, the sum stays above unity. This shows how feature extraction works. Note that HMM-DYN has $\hat{p}_{11} + \hat{p}_{22} = 1.61$, making it poor in signalling turning points. The out-of-sample forecast errors of HMM-DYN become only moderately larger than for MBC, as measured by MSE, but ECE is dramatically higher than for MBC, as expected according to Theorems 1 and 2.

5. CONCLUSIONS

It is suprising that so little attention has been paid to the practical use of probabilistic turning point forecasts. The starting point of this study was to find a device that produces accurate turning point *inference*. We introduced a new way to use HMM as a classifier, emphasising the inference aspect, where the cost function contains an extra penalty for wrong decisions. The resulting procedure provides a simple way to utilize leading information in a vector series.

We recommend applying well known classification error bounds that report the maximal risk of wrong inference and allow for comparing methods according to this criterion. We presented some theoretical results showing why MBC works where earlier uses of HMM have failed. The ML approach leads to high Markov probabilities of staying in the current regime, especially in an expansion. As a result, a recession will often be signalled late. MBC is designed to work precisely at turning points.

The method was used on Swedish and US data. MBC was compared with several other models based on HMM and a naive forecast. MBC beats all the alternatives and filtering proved to be of great benefit. Reliable inference for Swedish data is obtained for lead time one quarter as compared to the publication date of production statistics. In the case of the US, MBC was first applied on univariate data. The forecast accuracy was much worse than in the case where two variables were included.

The data-driven procedure presented here was developed as a solution to a practical problem of forecasting business cycle turning points, and is now regularly applied in forecasting the Swedish economy. The method uses all available information, including knowledge of past regimes, but it is required only to produce dichotomous inference.

17

This simplicity could prove useful also in other areas of application.

ACKNOWLEDGEMENTS

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.

Birchenhall, C., Jessen, H., Osborn D., and Simpson, P. (1999), " Predicting U.S. Business-Cycle Regimes," *Journal of Business Economic Statistics*, 17, 313-323.

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

Clements, M.P. and Hendry, D.F.: *Forecasting Non-Stationary Economic Time Series*, Zenthen Lecture Book Series, Ed.: K.G. Persson, The MIT Presss, Cambridge, Mass., 1999.

Cox, D.R. (1961), "Prediction by Exponentially Weighted Moving Averages and Related Methods," *Journal of the Royal Statistical Society*, B, 23, 414-422.

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.

Estrella, A, and Mishkin F.S. (1998), "Predicting US Recessions: Financial Variables as Leading Indicators," *The Review of Economics and Statistics*, 1, 45-61.

Filardo, A. (1994), "Business-Cycle Phases and Their transition Dynamics", *Journal of Business Economics Statistics*, 12, 299-308.

Fintzen, D., and Stekler H.O. (1999), "Why did forecasters fail to predict the 1990 recession", *International Journal of Forecasting*, 15, 309-323.

Fukanaga, K. (1972), *Introduction to Statistical Pattern Recognition*, Academic Press.

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

Ivanova, D., Lahiri, K., and Seitz F. (2000), "Interest rate spreads as predictors of German inflation and business cycle," *International Journal of Forecasting*, 39-58.

Keilis-Borok, V., Stock, J.H., Soloviev, A., and Mikhalev, P. (2000), "Pre-recession Pattern of Six Economic Indicator in the USA", *Journal of Forecasting*, 19, 65-80.

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.

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.

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

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 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.

Swanson, N.R. and White, H. (1997), "Forecasting Economic Time Series Using Flexible Versus Fixed Specification and Linear Versus Nonlinear Econometric Models", *International Journal of Forecasting*, 13, 439-461.

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

Table I Forecast accuracy of models for Sweden's turning points (lead $l = 1$).

Model	Smooth. par. λ_1	λ_2	λ_3	Markov prob. $\hat{p}_{11}$	$\hat{p}_{22}$	In-Sample MSE	ECE	S	Out-of-S. MSE	ECE	S
MBC	.2	.7	.1	.93	.47	.020	.000	.013	.007	.000	.005
MBC-UF	-	-	-	.85	.56	.061	.054	.059	.044	.059	.049
HMM-STAT	-	-	-	.68	.32	.064	.081	.069	.072	.147	.097
HMM-DYN	-	-	-	.92	.83	.072	.068	.071	.026	.059	.037
N	-	-	-	-	-	.218	.320	.252	.242	.323	.269

Legend: The parameter λ is the weight of the last observation in EWMA. The score S is $2/3 \times MSE + 1/3 \times ECE$.

Table II Forecast accuracy of univariate models for US turning points (lead $l = 0, 1, 2$).

Model	Smooth. par. λ	Markov prob. $\hat{p}_{11}$	$\hat{p}_{22}$	In-Sample MSE	ECE	S	Out-of-S. MSE	ECE	S
MBC $(l=0)$	.2	.81	.20	.062	.074	.066	.067	.100	.128
HMM	-	.92	.75	.132	.173	.146	.123	.138	.128
MBC $(l=1)$	.5	.87	.04	.068	.075	.070	.074	.101	.083
HMM	-	.92	.75	.107	.150	.121	.088	.089	.088
MBC $(l=2)$	.6	.84	.08	.080	.114	.091	.096	.141	.111
HMM	-	.92	.75	.117	.165	.133	.119	.141	.126

Table III Forecast accuracy of bivariate and naive models for US turning points (lead $l = 0, 1, 2$).

Model	Smooth. par. λ_1	λ_2	Markov prob. $\hat{p}_{11}$	$\hat{p}_{22}$	In-Sample MSE	ECE	S	Out-of-S. MSE	ECE	S
MBC	.7	.5	.90	.26	.032	.025	.030	.032	.038	.034
MBC-UF $(l=0)$	-	-	.97	.76	.034	.037	.035	.043	.050	.045
HMM-STAT	-	-	.85	.15	.051	.086	.063	.040	.038	.039
HMM-DYN	-	-	.94	.67	.035	.037	.036	.035	.063	.044
N	-	-	-	-	.126	.148	.133	.136	.163	.145
MBC	.2	1.0	.90	.05	.068	.075	.070	.060	.076	.065
MBC-UF $(l=1)$	-	-	.91	.26	.076	.088	.080	.076	.101	.084
HMM-STAT	-	-	.85	.15	.079	.088	.082	.076	.101	.084
HMM-DYN	-	-	.94	.67	.076	.113	.088	.084	.114	.094
N	-	-	-	-	.126	.148	.133	.136	.163	.145
MBC	.1	.5	.82	.05	.078	.114	.090	.109	.115	.111
MBC-UF $(l=2)$	-	-	.90	.15	.082	.114	.093	.104	.192	.133
HMM-STAT	-	-	.85	.15	.082	.165	.110	.103	.180	.129
HMM-DYN	-	-	.94	.67	.105	.177	.129	.115	.141	.124
N	-	-	-	-	.126	.148	.133	.136	.163	.145