INNOCENZO GASPARINI INSTITUTE FOR ECONOMIC RESEARCH

Institutional Members: CEPR, NBER and Università Bocconi

WORKING PAPER SERIES

Leading Indicators: What Have We Learned?

Massimiliano Marcellino

Working Paper n. 286

March 2005

IGIER – Università Bocconi, Via Salasco 5, 20136 Milano –Italy
http://www.igier.uni-bocconi.it

Leading Indicators: What Have We Learned?

Massimiliano Marcellino*
IEP-Bocconi University, IGIER and CEPR
massimiliano.marcellino@uni-bocconi.it

Abstract

We provide a summary updated guide for the construction, use and evaluation of leading indicators, and an assessment of the most relevant recent developments in this field of economic forecasting. To begin with, we analyze the problem of selecting a target coincident variable for the leading indicators, which requires coincident indicator selection, construction of composite coincident indexes, choice of filtering methods, and business cycle dating procedures to transform the continous target into a binary expansion/recession indicator. Next, we deal with criteria for choosing good leading indicators, and simple non-model based methods to combine them into composite indexes. Then, we examine models and methods to transform the leading indicators into forecasts of the target variable. Finally, we consider the evaluation of the resulting leading indicator based forecasts, and review the recent literature on the forecasting performance of leading indicators.

Keywords: Business Cycles, Leading Indicators, Coincident Indicators, Turning Points, Forecasting
J.E.L. Classification: E32, E37, C53

*In preparation for Elliot, G., Granger, C.W.J. and Timmermann, A. (eds), *Handbook of Economic Forecasting*, Elsevier - North Holland. I am grateful to participants at the Rady Conference on Economic Forecasting (UCSD) and to seminar participants at the Dutch Central Bank, University of Barcelona and Università Tor Vergata for useful comments on a previous draft; to Maximo Camacho, Jim Hamilton, Chang-Jin Kim, Charles Nelson, and Gabriel Perez-Quiros for sharing their code; to Ataman Ozyildirim at the Conference Board and Anirvan Banerji at ECRI for data and information; and to Andrea Carriero and Alice Ghezzi for excellent research assistance.

Contents

1 Introduction

Since the pioneering work of Mitchell and Burns (1938) and Burns and Mitchell (1946), leading indicators have attracted considerable attention, in particular by politicians and business people, who consider them as a useful tool for predicting future economic conditions. Economists and econometricians have developed more mixed feelings towards the leading indicators, starting with Koopmans's (1947) critique of the work of Burns and Mitchell, considered as an exercise in "measurement without theory". The resulting debate has stimulated the production of a vast literature that deals with the different aspects of the leading indicators, ranging from the choice and evaluation of the best indicators, possibly combined in composite indexes, to the development of more and more sophisticated methods to relate them to the target variable.

In this chapter we wish to provide a summary updated guide for the construction, use and evaluation of leading indicators and, more important, an assessment of the most relevant recent developments in this field of economic forecasting, starting with a series of contributions that were published in 1989 and have substantially changed the way to construct, use and evaluate leading indicators.

Stock and Watson (1989, SW) improved in five main respects the by then current practice in indicator analysis. First, they formalized Burns and Mitchell's (1946) notion that business cycles represent co-movements in a set of series by estimating a coincident index of economic activity as the unobservable factor in a dynamic factor model for four coincident indicators, industrial production, real disposable income, hours of work and sales. Similar variables were tracked by the members of the NBER's business cycle dating committee, and were combined with equal weights, after a proper standardization, in the Department of Commerce composite coincident indicator, now produced by the Conference Board, CCI_{CB}. In practice, Stock and Watson's index, CCI_{SW}, turns out to be very similar to the CCI_{CB}, but the former is derived in a formal framework that provides a sounder statistical basis for its construction and evaluation.

Their second important contribution is in the selection of candidate leading indicators. The latter were typically chosen on the basis of correlation analysis, spectral coherence, or the lead-time in their turning points with respect to the target variable. Stock and Watson conducted a systematic regression based analysis, starting with a large set of indicators, whose major benefit is that indicators are not evaluated in isolation but as a group. Perhaps the most interesting result in this part of their research is that several financial indicators turned out to have good leading properties. Burns and Mitchell themselves stressed the potential usefulness of stock price indexes, but the term structure of interest rates was rarely used until Stock and Watson pointed out its relevance.

The third contribution of Stock and Watson (1989) is to jointly model the coincident and the leading indicators. This stresses the importance of a clear definition of the target variable and its relationship with the leading indicator. In particular, their composite leading indicator, CLI_{SW}, aims at predicting the future 6-month growth in the CCI_{SW}. This is for example rather different from the goal of the Conference Board leading index, CLI_{CB}, which is to anticipate turning points in the level of the CCI_{CB}.

The fourth contribution is to cast the problem in a state space framework, which allows the joint resolution of a set of data problems such as the identification and removal of outliers, the treatment of data revisions, and the use of non-timely indicators whose most recent unavailable data can be substituted with forecasts. The seasonal adjustment of seasonal series could be also included in the framework, though Stock and Watson worked with seasonally adjusted data.

The final main contribution of Stock and Watson (1989) is to develop an index of leading indicators that produces early warnings of recession, CRI_{SW}, in the form of a probability that a recession will take place in the next six months. This required two main innovations. First, after a careful discussion of how to formally define a recession, they developed a pattern recognition algorithm that classifies future forecasted paths of the coincident indicator as belonging to a recessionary phase or not. Second, they introduced simulation methods to compute these forecasts of the coincident indicator and transform them into a probability of recession forecast. Though simulation methods were used earlier in the leading indicator literature, e.g. Wecker (1979) and Kling (1987), their use became widespread only in the '90s.

These five points were further developed in Stock and Watson (1991, 1992), while the performance of the CLI_{SW} and CRI_{SW} in the recessions of the early 90's and beginning of this century was evaluated in Stock and Watson (1993, 2003b). Unfortunately, the indicators did not perform particularly well in these two episodes, as most other forecasts. The main reason is that the leading characteristics of the variables change over time, and the financial indicators that worked very well in the '80s and got relevant weights in the CLI_{SW} were not particularly useful in the two subsequent recessions. We will discuss in more details this issue in the text, since it is the biggest problem and challenge for leading indicator forecasts. But the theoretical contributions of SW retain their importance, as testified by the substantial amount of additional research they have generated. The latter is reviewed in the chapter, together with suggestions for further improvements.

The main criticism Sims (1989) raised in his comment to Stock and Watson (1989) is the use of a constant parameter statistical model (estimated with classical rather than Bayesian methods). This belongs to the old debate on the characterization of business cycles as extrinsic phenomena, i.e. generated by the arrival of external shocks propagated through a liner model, versus intrinsic phenomena, i.e., generated by the nonlinear development of the endogenous variables. The main problem with the latter view, at least implicitly supported also by Burns and Mitchell that treated expansions and recessions as two different periods, was the difficulty of casting it into a simple and testable statistical framework, an issue addressed, again in 1989, by Hamilton.

Hamilton (1989) proposed a model were the mean of a random variable evolves according to an unobservable two-state Markov process, an idea developed in a different and simpler context also by Neftci (1982). Hamilton's main contribution to the business cycle literature, and in particular to that part related to leading indicators, is threefold. First, using the available data on one or more coincident series it is possible to infer the probability of being in an expansion or in a recession. Therefore, it is possible to substitute the judgemental dating of the business cycle, and the resulting categorization of observations into expansions and recessions, with a proper statistical method, which in addition can be easily implemented in real time.

Second, it is possible to jointly model coincident and leading indicators, since the latter should be driven by the same Markov process but with a lead. This intuition is formalized in Hamilton and Perez-Quiros (1996), and in a set of other studies reviewed below, who use the CLI_{CB} and other leading indicators to improve forecasts of the future status of the economy. Similarly, the method could be used for indicator selection or for the construction of composite indicators, along the lines suggested in a linear framework by SW.

Third, and related to the previous point, the method can be easily used to produce point or probability forecasts of the coincident variable, and also analytical forecasts of the probability of being in recession in or within a certain future date.

As in the case of Stock and Watson (1989), and perhaps even more, Hamilton (1989) has generated an impressive amount of subsequent research, with the part more closely related to business cycle analysis, and in particular to leading indicators, reviewed below. Here it is worth mentioning the work by Diebold and Rudebusch (1996), which allows the parameters of the model for the factor in Stock and Watson (1989) to change over the business cycle according to a Markov process. Kim and Nelson (1998) estimated the same model but in a Bayesian framework using the Gibbs sampler, as detailed below, therefore addressing both of Sims' criticisms reported above. Unfortunately, both papers confine themselves to the construction of a coincident indicator and do not consider the issue of leading indicators.

The third very relevant contribution to the leading indicator literature published in 1989 is Diebold and Rudebusch (1989). They used an extension of Neftci's (1982) rule to transform the CLI_{CB} into probabilities of recession, and compared them with the actual NBER peak and trough dates using a set of scoring rules. Among them, the quadratic probability score, which is the counterpart of the mean square forecast error in the probability forecast context, is by now currently adopted in the literature. In two follow-up papers, Diebold and Rudebusch (1991a,1991b), they repeated the evaluation exercise for, respectively, point and probability forecasts, but using a real time dataset rather than the latest available time series for the CLI_{CB}. The results were striking and pointed to a substantial deterioration in the predictive ability of the CLI_{CB}. The lessons are that real time data are particularly important for the evaluation of leading indicators since often revisions are driven by the past performance of the indicators, even though this practice has been substantially reduced in the '90s, and that indicators subject to only minor or no revisions, e.g. financial variables, can be more easily scored.

Perhaps as a consequence of the focus of Diebold and Rudebusch (1989) on the prediction of 0/1 events, a large literature was developed in the '90s on the use of binary models to predict business cycle phases, e.g., Estrella and Mishkin (1998) and Birchenhall et al. (1999). The basic logit or probit specifications have been also extended in a variety of ways, and a comprehensive survey of contributions in this area is presented below.

The chapter is organized as follows. In Section 2 we analyze the problem of selecting a target coincident variable for the leading indicators, which can be split into coincident indicator selection, construction of composite coincident indexes, choice of filtering methods, and business cycle dating procedures to transform the continuos target into a binary expansion/recession indicator. In section 3 we deal with criteria for choosing good leading indicators, and simple non model based methods to combine them into composite indexes. In section 4 we examine models and methods to transform the leading indicators into forecasts of the target variable, possibly after grouping the single leading indicators into a composite index. In Section 5 we consider the evaluation of the resulting leading indicator based forecasts, discuss several empirical examples, and review the recent literature on the forecasting performance of leading indicators. Finally, in Section 6 we summarize what we have learned about leading indicators from the recent literature and suggest directions for further research.

2 Selection of the target variable

The starting point for the construction of leading indicators is the choice of the target variable, namely, the variable that the indicators are supposed to lead. Such a choice is discussed in the first subsection. The second subsection analyzes a set of variable transformations and dating procedures, which are relevant, respectively, to emphasize the cyclical properties of the target variable and to transform it into a binary expansion/recession indicator. The final subsection presents some empirical examples as an illustration of the theoretical material.

2.1 Choice of variable

Burns and Mitchell (1946, p. 3) proposed that:

> "...a cycle consists of expansions occurring at about the same time in many economic activities...."

Yet, later on in the same book (p. 72) they stated:

> "Aggregate activity can be given a definite meaning and made conceptually measurable by identifying it with gross national product."

These quotes underlie the two most common choices of target variable: either a single indicator that is closely related to GDP but available at the monthly level, a case considered in the first sub-subsection, or a composite index of coincident indicators, as detailed in the second sub-subsection.

2.1.1 Single variables

As mentioned above, GDP could provide a reliable summary of the current economic conditions if it were available on a monthly basis. Though both in the US and in Europe there is a growing interest in increasing the sampling frequency of GDP from quarterly to monthly, the current results are still too preliminary to rely on them.

In the past, industrial production provided a good proxy for the fluctuations of GDP, and it is still currently monitored for example by the NBER business cycle dating committee and by the Conference Board in the US, in conjunction with other indicators. Yet, the ever rising share of services compared with the manufacturing, mining and gas and electric utility industries casts more and more doubts on the usefulness of IP as a single coincident indicator.

Another common indicator is the volume of sales of the manufacturing, wholesale and retail sectors, adjusted for price changes so as to proxy real total spending. Its main drawback, as in the case of IP, is its partial coverage of the economy.

A variable with a close to global coverage is real personal income less transfers, that underlies consumption decisions and aggregate spending. Yet, unusual productivity growth and favorable terms of trade can make income behave differently from payroll employment, the other most common indicator with economy wide coverage. More precisely, the monitored series is usually the number of employees on non-agricultural payrolls, whose changes reflect the net hiring (both permanent and transitory) and firing in the whole economy, with the exception of the smallest businesses and the agricultural sector.

Some authors focused on unemployment rather than employment, e.g. Boldin (1994) or Chin, Geweke and Miller (2000), on the grounds that the series is timely available and subject to minor revisions. Yet, typically unemployment is slightly lagging and not coincident.

Overall, it is difficult to identify a single variable that provides a good measure of current economic conditions and is available on a monthly basis. Therefore, it is preferable to consider combinations of several coincident indicators.

2.1.2 Composite indexes

The monitoring of several coincident indicators can be done either informally, for example the NBER business cycle dating committee examines the joint evolution of IP, employment, sales and real disposable income, see e.g. Hall et al. (2003), or formally, by combining the indicators into a composite index. A composite coincident index can be constructed in a non model based or in a model based framework, and we now review the main approaches within each category.

In the non model based framework, the coincident indicators are first transformed to have similar ranges, and then aggregated using equal or different weights. A clear illustration is provided by (a slightly simplified version of) the step-wise procedure implemented by the Conference Board, CB (previously by the Department of Commerce, DOC), see www.globalindicators.com for details.

First, for each individual indicator, x_{it}, month-to-month symmetric percentage changes (spc) are computed as $x_{it\_spc} = 200 * (x_{it} - x_{it-1})/(x_{it} + x_{it+1})$. Second, for each x_{it_spc} a volatility measure, v_i, is computed as the inverse of its standard deviation. Third, each x_{it_spc} is adjusted to equalize the volatility of the components, the standardization factor being $s_i = v_i / \sum_i v_i$. Fourth, the standardized components, $m_{it} = s_i x_{it\_spc}$, are summed together with equal weights, yielding $m_t = \sum_i m_{it}$. Fifth, the index in levels is computed as

$$CCI_t = CCI_{t-1} * (200 + m_t)/(200 - m_t) \qquad (1)$$

with the starting condition

$$CCI_1 = (200 + m_1)/(200 - m_1).$$

Finally, rebasing CCI to average 100 in 1996 yields the CCI_{CB}.

Within the model based approaches, two main methodologies have emerged: dynamic factor models and Markov switching models. In both cases there is a single unobservable force underlying the current status of the economy, but in the former approach this is a continuous variable, while in the latter it is a discrete variable that evolves according to a Markov chain. We now review these two methodologies, highlighting their pros and cons.

Dynamic factor models were developed by Geweke (1977) and Sargent and Sims (1977), but their use became well known to most business cycle analysts after the publication of

Stock and Watson's (1989, SW) attempt to provide a formal probabilistic basis for Burns and Mitchell's coincident and leading indicators, with subsequent refinements of the methodology in Stock and Watson (1991, 1992). The rationale of the approach is that a set of variables is driven by a limited number of common forces and by idiosyncratic components that are either uncorrelated across the variables under analysis or in any case common to only a limited subset of them. The particular model that SW adopted is the following,

$$\Delta x_t = \beta + \gamma(L)\Delta C_t + u_t \tag{2}$$
$$D(L)u_t = e_t \tag{3}$$
$$\phi(L)\Delta C_t = \delta + v_t \tag{4}$$

where x_t includes the (logs of the) four coincident variables used by the CB for their CCI_{CB}, the only difference being the use of hours of work instead of employment since the former provides a more direct measure of fluctuations in labor input. C_t is the single factor driving all variables, while u_t is the idiosyncratic component; Δ indicates the first difference operator, L is the lag operator and $\gamma(L)$, $D(L)$, $\phi(L)$ are, respectively, vector, matrix and scalar lag polynomials. SW used first differenced variables since unit root tests indicated that the coincident indexes were integrated, but not cointegrated. The model is identified by assuming that $D(L)$ is diagonal and e_t and v_t are mutually and serially uncorrelated at all leads and lags, which ensures that the common and the idiosyncratic components are uncorrelated. Moreover, ΔC_t should affect contemporaneously at least one coincident variable. Notice that the hypothesis of one factor, ΔC_t, does not mean that there is a unique source of aggregate fluctuations, but rather that different shocks have proportional dynamic effects on the variables.

For estimation, the model in (2)-(4) is augmented by the identity

$$C_{t-1} = \Delta C_{t-1} + C_{t-2,} \tag{5}$$

and cast into state-space form. The Kalman filter can then be used to write down the likelihood function, which is in turn maximized to obtain parameter and factor estimates, all the details are presented in Stock and Watson (1991).

A few additional comments are in order. First, the composite coincident index, CCI_{SWt}, is obtained through the Kalman filter as the minimum mean squared error linear estimator of C_t using information on the coincident variables up to period t. Hence, the procedure can be implemented in real time, conditional on the availability of data on the coincident variables. By using the Kalman smoother rather than the filter, it is possible to obtain end of period estimates of the state of the economy, i.e., $C_{t|T}$. Second, it is possible to obtain a direct measure of the contribution of each coincident indicator in x_t to the index by computing the response of the latter to a unit impulse in the former. Third, since data on some coincident indicator are published with delay, they can be treated as missing observations and estimated within the state-space framework. Moreover, the possibility of measurement error in the first releases of the coincident indicators can be also taken into consideration by adding an error term to the measurement equation. This is an important feature since data revisions are frequent and can be substantial, as for example testified by the revised US GDP growth rate data for 2001. Fourth, a particular time varying pattern in the parameters of the lag polynomials $D(L)$ and $\phi(L)$ can be allowed by using a time-varying transition matrix. Fifth, standard errors around the coincident index can be computed, even though they were not reported by SW.

The cyclical structure of CCI_{SW} closely follows the NBER expansions and recessions, and the correlation of two quarters growth rates in CCI_{SW} and real GDP was about .86 over the period 1959-87. Stock and Watson (1991) also compared their CCI_{SW} with the DOC's one, finding that the overall relative importance of the single indicators is roughly similar (but the weights are different since the latter index is made up of contemporaneous indicators only), the correlation of the levels of the composite indexes was close to 0.94, again over the period 1959-87, and the coherence of their growth rates at business cycle frequency was even higher.

These findings provide support for the simple averaging methodology originated at the NBER and then further developed at the DOC and the CB, but they also question the practical usefulness of the SW's approach, which is substantially more complicated. Overall, the SW methodology, and more generally model based index construction, are worth their cost since they provide a proper statistical framework that, for example, permits the computation of standard errors around the composite index, the unified treatment of data revisions and missing observations, the possibility of using time-varying parameters and, as we will see in more details in Section 4, a coherent framework for the development of composite leading indexes.

A possible drawback of SW's procedure is that it requires an ex-ante classification of variables into coincident and leading or lagging, even though this is common practice in this literature, and it cannot be directly extended to analyze large datasets because of computational problems, see Section 4.2 for details. Forni, Hallin, Lippi and Reichlin (2000, 2001 FHLR henceforth) proposed an alternative factor based methodology that addresses both issues, and applied it to the derivation of a composite coincident indicator for the Euro area. They analyzed a large set of macroeconomic time series for each country of the Euro area using a dynamic factor model, and decomposed each time series into a common and an idiosyncratic component, where the former is the part of the variable explained by common Euro area shocks, the latter by variable specific shocks. The CCI_{FHLR} is obtained as a weighted average of the common components of the interpolated monthly GDP series for each country, where the weights are proportional to GDP, and takes into account both within and across-countries cross correlations.

More specifically, the model FHLR adopted is

$$x_{it} = b_i^{'}(L)v_t + \xi_{it}, \quad i = 1, ..., N, \quad t = 1, ..., T, \tag{6}$$

where x_{it} is a stationary univariate random variable, v_t is a $q \times 1$ vector of common shocks, $\chi_{it} = x_{it} - \xi_{it}$ is the common component of x_{it}, and ξ_{it} is its idiosyncratic component. The shock v_t is an orthonormal white noise process, so that $var(v_{jt}) = 1$, $cov(v_t, v_{t-k}) = 0$, and $cov(v_{jt}, v_{st-k}) = 0$ for any $j \neq s$, t and k. $\xi_N = \{\xi_{1t}, ..., \xi_{Nt}\}^{'}$ is a wide sense stationary process, and $cov(\xi_{jt}, v_{st-k}) = 0$ for any j, s, t and k. $b_i(L)$ is a $q \times 1$ vector of square summable, bilateral filters, for any i. Notice that SW's factor model (2) is obtained as a particular case of (6) when there is one common shock ($q = 1$), $b_i(L) = \gamma_i(L)/\phi(L)$, and the idiosyncratic components are assumed to be orthogonal.

Grouping the variables into $x_{Nt} = \{x_{1t}, ..., x_{Nt}\}^{'}$, FHLR also required x_{Nt} (and χ_{Nt}, ξ_{Nt} that are similarly defined) to have rational spectral density matrices, Σ_N^x, Σ_N^χ, and Σ_N^ξ, respectively. To achieve identification, they assumed that the first (largest) idiosyncratic dynamic eigenvalue, λ_{N1}^{ξ}, is uniformly bounded, and that the first (largest) q common dynamic eigenvalues, $\lambda_{N1}^{\chi}, ..., \lambda_{Nq}^{\chi}$, diverge, where dynamic eigenvalues are the eigenvalues of the spectral density matrix, see e.g. Brillinger (1981, Chap. 9). In words, the former condition limits the effects of ξ_{it} on other cross-sectional units. The latter, instead, requires v_t to affect infinitely many units.

Let us assume for the moment that the number of common shocks is known. Then, FHLR suggested to estimate the common component of χ_{it} with the following step-wise procedure.

(i) Estimate the spectral density matrix of x_{Nt} as

$$\Sigma_N^T(\theta_h) = \sum_{k=-M}^{M} \Gamma_{Nk}^T \omega_k e^{-ik\theta_h}, \quad \theta_h = 2\pi h/(2M+1), \quad h = 0, ..., 2M, \tag{7}$$

where Γ_{Nk}^T is the sample covariance matrix of x_{Nt} and x_{Nt-k}, ω_k is the Bartlett lag window of size M ($\omega_k = 1 - k/(M+1)$), and M diverges but M/T tends to zero.

(ii) Calculate the first q eigenvectors of $\Sigma_N^T(\theta_h)$, $p_{Nj}^T(\theta_h)$, and the associated eigenvalues $\lambda_{j\theta}^x$, $j = 1, ..., q$, for $h = 0, ..., 2M$.

(iii) Define $p_{Nj}^T(L)$ as

$$p_{Nj}^T(L) = \sum_{k=-M}^{M} p_{Nj,k}^T L^k, \quad p_{Nj,k}^T = \frac{1}{2M+1} \sum_{h=0}^{2M} p_{Nj}^T(\theta_h) e^{ik\theta_h}, \quad k = -M, ..., M. \tag{8}$$

$p_{Nj}^T(L)x_{Nt}$, $j = 1, .., q$, are the first q estimated dynamic principal components of x_{Nt}.

(iv) The estimated common component of x_{it}, $\widehat{\chi}_{it}$, is the projection of x_{it} on present, past, and future dynamic principal components. FHLR proved that, under mild conditions, $\widehat{\chi}_{it}$ is a consistent estimator of χ_{it} when N and T diverge. Once the common component is estimated, the idiosyncratic one is obtained simply as a residual, namely, $\widehat{\xi}_{it} = x_{it} - \widehat{\chi}_{it}$.

In practice, M and the number of leads (s) and lags (g) of $p_{Nj}^T(L)x_{Nt}$ to be included in the projection can be either chosen a priori or determined by minimizing the information criterion

$$\frac{T}{N} \sum_{i=1}^{N} \log \widehat{\sigma}_i + 2q(g+s+1), \tag{9}$$

where $\widehat{\sigma}_i$ is the estimated variance of $\widehat{\xi}_{it}$. Finally, FHLR suggested to determine the number of factors, q, on the basis of two properties: (a) the average over frequencies of the first q dynamic eigenvalues diverges, while the average of the $q+1^{th}$ does not; (b) there should be a big gap between the variance of x_{Nt} explained by the first q dynamic principal components and that explained by the $q+1^{th}$ principal component. An information criterion could be also used. In particular, the criterion that FHLR suggested for selection of g and s, equation (9) above, could be minimized also with respect to q.

The methodology was further refined by Altissimo et al. (2001) for real time implementation, and it is currently adopted to produce the CEPR's composite coincident indicator for the euro area, Eurocoin (see www.cepr.org). In particular, they exploited the large cross-sectional dimension for forecasting indicators available with delay and for filtering out high frequency dynamics. An alternative coincident index for the Euro area following the SW methodology was proposed by Proietti and Moauro (2003).

As mentioned in the Introduction, the main criticism Sims (1989) raised in his comment to Stock and Watson (1989) is the use of a constant parameter model (even though, as remarked above, their framework is flexible enough to allow for parameter variation), and a similar critique can be addressed to the FHLR's method. Hamilton's (1989) Markov switching model is a powerful response to this criticism, since it allows the growth rate of the variables (and possibly their dynamics) to depend on the status of the business cycle. A basic version of the model can be written as

$$\Delta x_t = c_{s_t} + A_{s_t} \Delta x_{t-1} + u_t, \tag{10}$$
$$u_t \sim i.i.d.N(0, \Sigma) \tag{11}$$

where, as in (2), x_t includes the coincident variables under analysis (or a single composite index), while s_t measures the status of the business cycle, with $s_t = 1$ in recessions and $s_t = 0$ in expansions, and both the deterministic component and the dynamics can change over different business cycle phases. The binary state variable s_t is not observable, but the values of the coincident indicators provide information on it.

With respect to the factor model based analysis, there is again a single unobservable force underlying the evolution of the indicators but, first, it is discrete rather than continuous and, second, it does not directly affect or summarize the variables but rather indirectly determines their behaviour that can change substantially over different phases of the cycle.

To close the model and estimate its parameters, an equation describing the behaviour of s_t is required, and it cannot be of autoregressive form as (4) since s_t is a binary variable. Hamilton (1989) proposed to adopt the Markov switching (MS) model, where

$$\Pr(s_t = j | s_{t-1} = i) = p_{ij}, \tag{12}$$

as previously considered by Lindgren (1978) and Neftci (1982) in simpler contexts. For expositional purposes we stick to the two states hypothesis, though there is some empirical evidence that three states can further improve the specification, representing recession, high growth and normal growth, see e.g. Kim and Murray (2002) for the US and Artis, Krolzig and Toro (2003) for the Euro area.

In our business cycle context, the quantity of special interest is an estimate of the unobservable current status of the economy and, assuming a mean square error loss function, the best estimator coincides with the conditional expectation of s_t given current and past information on x_t, which in turn is equivalent to the conditional probability

$$\zeta_{t|t} = \begin{pmatrix} \Pr(s_t = 0 | x_t, x_{t-1}, ..., x_1) \\ \Pr(s_t = 1 | x_t, x_{t-1}, ..., x_1) \end{pmatrix}. \tag{13}$$

Using simple probability rules, it follows that

$$\zeta_{t|t} = \begin{pmatrix} \frac{f(x_t | s_t=0, x_{t-1}, ..., x_1) \Pr(s_t=0 | x_{t-1}, ..., x_1)}{f(x_t | x_{t-1}, ..., x_1)} \\ \frac{f(x_t | s_t=1, x_{t-1}, ..., x_1) \Pr(s_t=1 | x_{t-1}, ..., x_1)}{f(x_t | x_{t-1}, ..., x_1)} \end{pmatrix}, \tag{14}$$

where

$$\Pr(s_t = i|x_{t-1}, ..., x_1) = \sum_{j=0}^{1} p_{ji} \Pr(s_{t-1} = j|x_{t-1}, ..., x_1), \tag{15}$$

$$f(x_t|s_t = i, x_{t-1}, ..., x_1) = \frac{1}{(2\pi)^{T/2}} |\Sigma|^{-1/2} \exp[-(\Delta x_t - c_i - A_i \Delta x_{t-1})' \Sigma^{-1} (\Delta x_t - c_i - A_i \Delta x_{t-1})/2],$$

$$f(x_t, s_t = i|x_{t-1}, ..., x_1) = f(x_t|s_t = i, x_{t-1}, ..., x_1) \Pr(s_t = i|x_{t-1}, ..., x_1),$$

$$f(x_t|x_{t-1}, ..., x_1) = \sum_{j=0}^{1} f(x_t, s_t = j|x_{t-1}, ..., x_1), \quad i = 0, 1.$$

Hamilton (1994) or Krolzig (1997) provide additional details on these computations, and formulae to calculate $\zeta_{t|T}$, i.e., the smoothed estimator of the probability of being in a given status in period t. Notice also that the first and last rows of (15) provide, respectively, the probability of the state and the density of the variables conditional on past information only, that will be used in Section 4 in a related context for forecasting.

For comparison and since it is rather common in empirical applications (see e.g. Neimira and Klein (1994) for the US and Artis et al. (1995) for the UK), it is useful to report Neftci's (1982) formula to compute the (posterior) probability of a turning point given the available data, as refined by Diebold and Rudebusch (1989). Defining

$$\Pi_t = \Pr(s_t = 1|x_t, , ..., x_1), \tag{16}$$

the formula is

$$\begin{aligned} \Pi_t &= \frac{A_1}{B_1 + C_1}, \\ A_1 &= (\Pi_{t-1} + p_{01}(1 - \Pi_{t-1}))f(x_t|s_t = 1, x_{t-1}, ..., x_1), \\ B_1 &= (\Pi_{t-1} + p_{01}(1 - \Pi_{t-1}))f(x_t|s_t = 1, x_{t-1}, ..., x_1), \\ C_1 &= (1 - \Pi_{t-1})(1 - p_{01})f(x_t|s_t = 0, x_{t-1}, ..., x_1). \end{aligned} \tag{17}$$

The corresponding second element of $\zeta_{t|t}$ in (14) can be written as

$$\begin{aligned} \Pi_t &= \frac{A_2}{B_2 + C_2}, \\ A_2 &= (\Pi_{t-1} - \Pi_{t-1}p_{01} + p_{01}(1 - \Pi_{t-1}))f(x_t|s_t = 1, x_{t-1}, ..., x_1), \\ B_2 &= (\Pi_{t-1} - \Pi_{t-1}p_{01} + p_{01}(1 - \Pi_{t-1}))f(x_t|s_t = 1, x_{t-1}, ..., x_1), \\ C_2 &= ((1 - \Pi_{t-1})(1 - p_{01}) + \Pi_{t-1}p_{01})f(x_t|s_t = 0, x_{t-1}, ..., x_1). \end{aligned} \tag{18}$$

Since in practice the probability of transition from expansion to recession, p_{01}, is very small (e.g., Diebold and Rudebusch (1989) set it at .02), the term $\Pi_{t-1}p_{01}$ is also very small and the two probabilities in (17) and (18) are very close. Yet, in general it is preferable to use the expression in (18) which is based on a more general model. Notice also that when $\Pi_t = 1$ the formula in (17) gives a constant value of 1 (e.g., Diebold and Rudebusch (1989) put an ad-hoc upper bound of .95 for the value that enters the recursive formula), while this does not happen with (18).

The model in (10)-(12) can be extended in several dimensions, for example to allow for more states and cointegration among the variables, see e.g. Krolzig, Marcellino and Mizon (2002), or time-varying probabilities, as e.g. in Diebold, Lee and Weinbach (1994) or Filardo (1994). The latter case is of special interest in our context when past values of the leading indicators, y, are the driving forces of the probabilities, as in Filardo (1994), who substituted (12) with

$$\Pr(s_t = i|s_{t-1} = j, x_{t-1}, ..., x_1, y_{t-1}, ..., y_1) = \frac{\exp(\theta y_{t-1})}{1 + \exp(\theta y_{t-1})}, \tag{19}$$

so that the first row of (15) should be modified into

$$\begin{aligned} &\Pr(s_t = i|x_{t-1}, ..., x_1) = \\ &= \frac{\exp(\theta y_{t-1})}{1 + \exp(\theta y_{t-1})} \Pr(s_{t-1} = j|x_{t-1}, ..., x_1) + \frac{1}{1 + \exp(\theta y_{t-1})} \Pr(s_{t-1} = i|x_{t-1}, ..., x_1). \end{aligned} \tag{20}$$

Another example is provided by Filardo and Gordon (1998), who used a probit model rather than a logistic specification for $\Pr(s_t = i|s_{t-1} = j, x_{t-1}, ..., x_1, y_{t-1}, ..., y_1)$, while Ravn and Sola (1999) warned against possible parameter instability of relationships such as (19). Raj (2002) provides a more detailed review of these and other extensions of the MS model.

As mentioned in the Introduction, factor models and Markov switching specifications capture two complementary and fundamental features of business cycles, namely, the diffusion of slow-down and recovery across many series and the different behavior of several indicators in expansions and recessions. They are not only flexible and powerful statistical tools but can be also given sound justifications from an economic theory point of view, see e.g. the overview in Diebold and Rudebusch (1996). The latter article represents also one of the earliest attempts to combine the two approaches, by allowing the factor underlying SW's model to evolve according to a Markov switching model. To provide support for their ideas, they fitted univariate and multivariate MS models to, respectively, the DOC's composite coincident indicator and its components, finding substantial evidence in favor of the MS specifications. Yet, they did not jointly estimate the factor MS model. Such a task was tackled by Chauvet (1998) and Kim and Yoo (1995), using an approximated maximum likelihood procedure developed by Kim (1994), and by Kim and Nelson (1998) and Filardo and Gordon (1999) using Gibbs sampler techniques introduced by Albert and Chib (1993a), Carter and Kohn (1994), and Shepard (1994).

In particular, Kim and Nelson (1998) substituted equation (4) in SW's model with

$$\begin{aligned} \phi(L)(\Delta C_t - \mu_{s_t} - \delta) &= v_t, \\ \mu_{s_t} &= \mu_0 + \mu_1 s_t, \end{aligned} \tag{21}$$

where the transition probabilities are either constant or follow a probit specification. They compared the (posterior) regime probabilities from the factor MS model estimated with the four SW's components with those from a univariate MS model for IP, concluding that the former are much more closely related with the NBER expansion/recession classification. Yet, such a result is not surprising since Filardo (1994) showed that time-varying probabilities are needed for the univariate MS model to provide a close match with the NBER classification. When the original SW's model is estimated using the Gibbs sampling approach, the posterior distributions of the parameters are very close to those obtained using (21) instead of (4), the main difference being a slightly larger persistence of the estimated factor. Filardo and Gordon (1999), focusing on the 1990 recession, also found a similar performance of the standard and MS factor model, while a multivariate MS model with time-varying probabilities performed best during the recessionary part of 1990 (but not significantly better in the remaining months). Finally, Kim and Nelson (1998) also found a close similarity of their composite coincident indicator and the equal weighted DOC's one, with correlation in the growth rates above .98.

Overall, no clear cut ranking of the multivariate model based approaches to coincident indicators construction emerges, but the resulting indexes are in general very similar, and close to the equal weighted ones. The positive aspect of this result is that estimation of the current economic condition is rather robust to the choice of method. Another implication is that pooling methods can be expected to yield no major improvements because of high correlation of all the indicators, but this is an issue that certainly deserves further investigation.

To conclude, notice that if the probability of the states is time varying, e.g. as in (19), and the indicators in y_t include a measure of the length of the current recession (or expansion), it is possible to allow and test for duration dependence, namely, for whether the current or past length of a business cycle phase influences its future duration. The test is based on the statistical significance of the parameter associated with the duration indicator in an equation such as (19). Earlier studies using non-parametric techniques, such as Diebold and Rudebusch (1990) or Diebold, Rudebusch and Sichel (1993), detected positive duration dependence for recessions but not for expansions. Such a finding was basically confirmed by Durland and McCurdy (1994) using a semi-Markov model with duration depending only on calendar time, by Filardo and Gordon (1998) in a univariate Markov switching framework that also relates duration to macroeconomic variables, and by Kim and Nelson (1998) in their multivariate factor MS model. Therefore, another interesting question to be addressed in Section 4 is whether leading indicators can be used to predict the duration of a business cycle phase.

2.2 Filtering and dating procedures

Once the choice of the measure of aggregate activity is made, two issues emerge: first the selection of the proper variable transformation, if any, and second the adoption of a business cycle dating rule that identifies the peaks and troughs in the series, and the associated expansionary and recessionary periods and their durations.

The choice of the variable transformation is related to the two broad definitions of the cycle recognized in the literature, the so-called classical cycle and the growth or deviation cycle. In the case of the deviation cycle, the focus is on the deviations of the rate of growth of the target variable from an appropriately defined trend rate of growth, while the classical cycle relies on the levels of the target variable.

Besides removing long term movements as in the deviation cycle, high frequency fluctuations can be also eliminated to obtain a filtered variable that satisfies the duration requirement in the original definition of Burns and Mitchell (1946, p.3):

"... in duration business cycles vary from more than one year to ten or twelve years; they are not divisible into shorter cycles of similar character with amplitudes approximating their own."

There is a large technical literature on methods of filtering the data. In line with the previous paragraph, Baxter and King (1999) argued that the ideal filter for cycle measurement must be customized to retain unaltered the amplitude of the business cycle periodic components, while removing high and low frequency components. This is known as a *band-pass* filter and, for example, when only cycles with frequency in the range 1.5-8 years is of interest, its theoretical frequency response function takes the rectangular form: $w(\omega) = \mathrm{I}(2\pi/(8s) \leq \omega \leq 2\pi/(1.5s))$, where $\mathrm{I}(\cdot)$ is the indicator function. Moreover, the phase displacement of the filter should always be zero, to preserve the timing of peaks and troughs; the latter requirement is satisfied by a symmetric filter.

Given the two business cycle frequencies, $\omega_{c1} = 2\pi/(8s)$ and $\omega_{c2} = 2\pi/(1.5s)$, the band-pass filter is

$$w_{bp}(L) = \frac{\omega_{c2} - \omega_{c1}}{\pi} + \sum_{j=1}^{\infty} \frac{\sin(\omega_{c2}j) - \sin(\omega_{c1}j)}{\pi j}(L^j + L^{-j}). \tag{22}$$

Thus, the ideal band-pass filter exists and is unique, but entails an infinite number of leads and lags, so in practice an approximation is required. Baxter and King (1999) showed that the K-terms approximation to the ideal filter (22) that is optimal in the sense of minimizing the integrated squared approximation error is simply (22) truncated at lag K. They proposed using a three years window, i.e. $K = 3s$, as a valid rule of thumb for macroeconomic time series. They also constrained the weights to sum up to zero, so that the resulting approximation is a detrending filter: denoting the truncated filter $w_{bp,K}(L) = w_0 + \sum_1^K w_j(L^j + L^{-j})$, the weights of the adjusted filter are $w_j - w_{bp,K}(1)/(2K+1)$, see e.g. Stock and Watson (1999a) for an application.

As an alternative, Christiano and Fitzgerald (1999) proposed to project the ideal filter on the available sample. If $c_t = w_{bp}(L)x_t$ denotes the ideal cyclical component, their proposal is to consider $\widehat{c}_t = E(c_t|x_1, ..., x_T)$, where x_t is given a parametric linear representation, e.g. an ARIMA model. They also found that for a wide class of macroeconomic time series the filter derived under the random walk assumption for x_t is feasible and handy.

Baxter and King (1999) did not consider the problem of estimating the cycle at the extremes of the available sample (the first and last three years), which is inconvenient for a real-time assessment of current business conditions. Christiano and Fitzgerald (1999) suggested to replace the out of sample missing observations by their best linear prediction under the random walk hypothesis. Yet, this can upweight the last and the first available observations. As a third alternative, Artis, Marcellino and Proietti (2003, AMP) designed a band-pass filter as the difference of two Hodrick Prescott (1997) detrending filters with parameters $\lambda = 1$ and $\lambda = 677.13$, where these values are selected to ensure that $\omega_{c1} = 2\pi/(8s)$ and $\omega_{c2} = 2\pi/(1.5s)$. The resulting estimates of the cycle are comparable to the Baxter and King cycle, although slightly noisier, without suffering from unavailability of the end of sample estimates

Working with growth rates of the coincident variables rather than levels, a convention typically adopted for the derivation of the composite indexes in the previous subsection, corresponds to the application of a filter whose theoretical frequency response function increases monotonically, starting at zero at the zero frequency. Therefore, growth cycles and deviation cycles need not be very similar.

In early post-war decades, especially in Western Europe, growth was relatively persistent and absolute declines in output were comparatively rare; the growth or deviation cycle then seemed to be of more analytical value, especially as inflexions in the rate of growth of output could reasonably be related to fluctuations in the levels of employment and unemployment. In more recent decades, however, there have been a number of instances of absolute decline in output, and popular description at any rate has focussed more on the classical cycle. The concern that de-trending methods can affect the information content of the series in unwanted ways, see e.g. Canova (1998), has reinforced the case for examining the classical cycle. The relationships among the three types of cycles are analyzed in more details below, after defining the dating algorithms.

In the U.S., the National Bureau of Economic Research (http://www.nber.org) provides a chronology of the classical business cycle since the early 1920s, based on the consensus of a set of coincident indicators concerning production, employment, real income and real sales, that is widely accepted among economists and policy makers, see e.g. Moore and Zarnowitz (1986). A similar chronology has been recently proposed for the euro area by the Center for Economic Policy Research (http://www.cepr.org).

Since the procedure underlying the NBER dating is informal and subject to substantial delays in the announcement of the peak and trough dates (which is rational to avoid later revisions), several alternative methods have been put forward and tested on the basis of their ability to closely reproduce the NBER classification.

The simplest approach, often followed by practitioners, is to identify a recession with at least two quarters of negative real GDP growth. Yet, the resulting chronology differs with respect to the NBER in a number of occasions, see e.g. Watson (1991) or Boldin (1994).

A more sophisticated procedure was developed by Bry and Boschan (1971) and further refined by Harding and Pagan (2001). In particular, for quarterly data on the log-difference of GDP or GNP (Δx_t), Harding and Pagan defined an expansion terminating sequence, ETS_t, and a recession terminating sequence, RTS_t, as follows:

$$\begin{array}{rcl} ETS_t & = & \{(\Delta x_{t+1} < 0) \cap (\Delta\Delta x_{t+2} < 0)\} \\ RTS_t & = & \{(\Delta x_{t+1} > 0) \cap (\Delta\Delta x_{t+2} > 0)\} \end{array} \tag{23}$$

The former defines a candidate point for a peak in the classical business cycle, which terminates the expansion, whereas the latter defines a candidate for a trough. When compared with the NBER dating, usually there are only minor discrepancies. Stock and Watson (1989) adopted an even more complicated rule for identifying peaks and troughs in their composite coincident index.

Within the Markov switching framework, a classification of the observations is automatically produced if the probability of being in a recession, $\Pr(s_t = 1|x_t, x_{t-1}, ..., x_1)$ in (13) or its smoothed version, is complemented with a threshold such that when $\Pr(s_t = 1|x_t, x_{t-1}, ..., x_1)$ is beyond the threshold the observation is marked as belonging to a recession. The turning points are then easily obtained as the dates of switching from expansion to recession, or vice versa. Among others, Boldin (1994) reported encouraging results using a MS model for unemployment, and Layton (1996) for the ECRI coincident index. Chauvet and Piger (2003) confirmed the positive results also with a real-time dataset and for a more up-to-date sample period.

Harding and Pagan (2004) compared their non-parametric rule with the MS approach, and further insight can be gained from Hamilton's (2004) comments on the paper and the authors' rejoinder. While the non-parametric rule produces simple, replicable and robust results, it lacks a sound economic justification and cannot be used for probabilistic statements on the current status of the economy. On the other hand, the MS model provides a general statistical framework to analyze business cycle phenomena, but the requirement of a parametric specification introduces a subjective element into the analysis and can necessitate a careful tailoring. Moreover, if the underlying model is linear, the MS recession indicator is not identified while pattern recognition works in any case.

AMP developed a dating algorithm based on the theory of Markov chains that retains the attractive features of the non-parametric methods, but allows the computation of the probability of being in a certain regime or of a phase switch. Moreover, the algorithm can be easily modified to introduce depth or amplitude restrictions, and to construct diffusion indices. Basically, the transition probabilities, p_{ij} in (12), are scored according to the pattern in the series x_t rather than within a parametric MS model. The resulting chronology for the euro area is very similar to the one proposed by the CEPR, and a similar result emerges for the US with respect to the NBER dating, with the exception of the last recession, see Table 3 and Section 2.3 below.

An alternative parametric procedure to compute the probability of being in a certain cyclical phase is to adopt a probit or logit model where the dependent variable is the NBER expansion/recession classification, and the regressors are the coincident indicators. For example, Birchenhall, Jessen, Osborn and Simpson (1999) showed that the fit of a logit model is very good in sample when the four NBER coincident indicators are used. They also found that the logit model outperformed a MS alternative, while Layton and Katsuura (2001) obtained the opposite ranking in a slightly different context.

The in-sample estimated parameters from the logit or probit models can be also used in combination with future available values of the coincident indicators to predict the future status of the economy, which is for example useful to conduct a real time dating exercise because of the mentioned delays in the NBER announcements.

So far, in agreement with most of the literature, we have classified observations into two phases, recessions and expansions, which are delimited by peaks and troughs in economic activity. However, multiphase characterizations of the business cycle are not lacking in the literature: the popular definition due to Burns and Mitchell (1946) postulated four states: expansion, recessions, contractions, recovery; see also Sichel (1994) for an ex-ante three phases characterization of the business cycle, Artis, Krolzig and Toro (2001) for an ex-post three-phases classification based on a model with Markov switching, and Layton and Katsuura (2001) for the use of multinomial logit models.

To conclude, having defined several alternative dating procedures, it is useful to return to the different notions of business cycle and recall a few basic facts about their dating, summarizing results in AMP.

First, neglecting duration ties, classical recessions (i.e. peak-trough dynamics in x_t), correspond to periods of prevailing negative growth, $\Delta x_t < 0$. In effect, negative growth is a sufficient, but not necessary, condition for a classical recession under the Bry and Boschan dating rule and later extensions. Periods of positive growth can be observed during a recession, provided that they are so short lived that they do not determine an exit from the recessionary state.

Second, turning points in x_t correspond to Δx_t crossing the zero line (from above zero if the turning point is a peak, from below in the presence of a trough in x_t). This is strictly true under the calculus rule, according to which $\Delta x_t < 0$ terminates the expansion.

Third, if x_t admits the log-additive decomposition, $x_t = \psi_t + \mu_t$, where ψ_t denotes the deviation cycle, then growth is in turn decomposed into cyclical and residual changes:

$$\Delta x_t = \Delta\psi_t + \Delta\mu_t.$$

Hence, assuming that $\Delta\mu_t$ is mostly due to growth in trend output, deviation cycle recessions correspond to periods of growth below potential growth, that is $\Delta x_t < \Delta\mu_t$. Using the same arguments, turning points correspond to Δx_t crossing $\Delta\mu_t$. When the sum of potential growth and cyclical growth is below zero, that is $\Delta\mu_t + \Delta\psi_t < 0$, a classical recession also occurs.

Finally, as an implication of the previous facts, classical recessions are always a subset of deviation cycle recessions, and there can be multiple classical recessionary episodes within a period of deviation cycle recessions. This suggests that an analysis of the deviation cycle can be more informative and relevant also from the economic policy point of view, even though more complicated because of the filtering issues related to the extraction of the deviation cycle.

2.3 Examples

In Figure 1 we graph four composite coincident indexes for the US: the Conference Board's equal weighted non model based CCI, the OECD coincident reference series which is a transformation of IP, the Stock and Watson's (1989) factor model based CCI, and the Kim and Nelson's (1998) bayesian MS factor model based CCI computed using the four coincident series combined in the CCI_{CB}. All indexes are normalized to have zero mean and unit standard deviation.

The Figure highlights the very similar behavior of all the CCIs, which in particular share the same pattern of peaks and troughs. The visual impression is confirmed by the correlations for the levels, and by those for the 6-month percentage changes reported in Table 1, the lowest value being 0.916 for CCI_{KN} and CCI_{OECD}. These values are in line with previous studies, see Section 2.1, and indicate that is possible to achieve a close to complete agreement on the status of the economy.

In Figure 2 we consider dating the US classical and deviation cycles. In the upper panel we graph the CCI_{CB} and the NBER expansion/recession classification. The figure highlights that the NBER recessions virtually coincide with the peak-trough periods in the CCI_{CB}. In the middle panel we graph the CCI_{CB} and the expansion/recession classification resulting from the AMP dating. The results are virtually identical with respect to the NBER (see also the first two columns of Table 3), with the noticeable difference that AMP identifies a double dip at the beginning of the new century with recessions in 2000:10-2001:12 and 2002:7-2003:4 versus 2001:3-2001:11 for the NBER. In the lower panel of Figure 2 we graph the HP band pass filtered CCI_{CB}, described in Section 2.2, and the AMP dating for the resulting deviation cycle. As discussed above, the classical cycle recessions are a subset of those for the deviation cycle, since the latter capture periods of lower growth even if not associated with declines in the level of the CCI.

Finally, in Figure 3 we report the (filtered) probability of recessions computed with two methods. In the upper panel we graph the probabilities resulting from the Kim and Nelson's (1998) bayesian MS factor model applied to the four coincident series combined in the CCI_{CB}. In the lower panel those from the AMP non-parametric MS approach applied to the CCI_{CB}. The results in the two panels are very similar, and the matching of peaks in these probabilities and NBER dated recessions is striking. The latter result supports the use of these methods for real-time dating of the business cycle. It is also worth noting that both methods attribute a probability close to 60% for a second short recession at the beginning of the century, in line with the AMP dating reported in the middle panel of Figure 2 but in contrast with the NBER dating.

3 Choice of the leading indicators

Once the target variable is identified, the leading indicators have to be selected. We discuss selection criteria in the first subsection; composite indexes in the second subsection; and a few examples in the final subsection.

3.1 Indicator selection

Since the pioneering work of Mitchell and Burns (1938), variable selection has rightly attracted considerable attention in the leading indicator literature, see e.g. Zarnowitz and Boschan (1975a,b) for a review of early procedures at the NBER and Department of Commerce. Moore and Shiskin (1967) formalized an often quoted scoring system (see e.g. Boehm (2001), Phillips (2003)), based mostly upon (i) consistent timing as a leading indicator (i.e., to systematically anticipate peaks and troughs, possibly with a rather constant lead time); (ii) conformity to the general business cycle (i.e., have good forecasting properties not only at peaks and troughs); (iii) economic significance (i.e., being supported by economic theory either as possible causes of business cycles or, perhaps more importantly, as quickly reacting to negative or positive shocks); (iv) statistical reliability of data collection (i.e., provide an accurate measure of the quantity of interest); (v) prompt availability without major later revisions (i.e., being timely and regularly available for an early evaluation of the expected economic conditions, without requiring subsequent modifications of the initial statements); (vi) smooth month to month changes (i.e., being free of major high frequency movements).

Some of these properties can be formally evaluated at different levels of sophistication. In particular, if a dating procedure is applied both to the coincident variable or index and to the candidate leading indicator, the resulting peak/trough dates can be compared and used to evaluate whether the peak structure of the leading indicator systematically anticipated that of the coincident indicator, with a stable lead time (property i)). An alternative procedure can be based on the statistical concordance of the binary expansion/recession indicators (resulting from the peak/trough dating) for the coincident and lagged leading variables, where the number of lags of the leading variable can be either fixed or chosen to maximize the concordance. A formal test for no concordance is defined below in Section 5.1. A third option is to run a logit/probit regression of the coincident expansion/recession binary indicator on the leading variable, evaluating the explanatory power of the latter. The major advantage of this procedure is that several leading indicators can be jointly considered to measure their partial contribution. Details on the implementation of this procedure are provided in Section 4.6.

To assess whether a leading indicator satisfies property ii), conformity to the general business cycle, it is preferable to consider it and the target coincident index as continuos variables rather than transforming them into binary indicators. Then, the set of available techniques includes frequency domain procedures (such as the spectral coherence and the phase lead), and several time domain methods, ranging from Granger causality tests in multivariate linear models, to the evaluation of the marginal predictive content of the leading indicators in sophisticated non-linear models, possibly with time varying parameters, see Section 4 for details on these methods. Within the time domain framework it is also possible to consider a set of additional relevant issues such as the presence of cointegration between the coincident and leading indicators, the determination of the number lags of the leading variable, or the significance of duration dependence. We defer a discussion of these topics to Section 4.

Property iii), economic significance, can be hardly formally measured, but it is quite important both to avoid the measurement without theory critique, e.g. Koopmans (1947), and to find indicators with stable leading characteristics. On the other hand, the lack of a commonly accepted theory of the origin of business cycles, see e.g. Fuhrer and Schuh (1998), makes it difficult to select a single indicator on the basis of its economic significance.

Properties iv) and v) have received considerable attention in the recent years and, together with economic theory developments, underlie the more and more widespread use of financial variables as leading indicators (due to their exact measurability, prompt availability and absence of revisions), combined with the adoption of real-time datasets for the assessment of the performance of the indicators, see Section 5 for details on these issues. Time delays in the availability of leading indicators are particularly problematic for the construction of composite leading indexes, and have been treated differently in the literature and in practice. Either preliminary values of the composite indexes are constructed excluding the unavailable indicators and later revised, along the tradition of the NBER and later of the Department of Commerce and the Conference Board, or the unavailable observations are substituted with forecasts, as in the factor based approaches described in Section 2.1 with reference to the composite coincident index. The latter solution is receiving increasing favor also within the traditional methodology, see e.g. McGuckin, Ozyildirim and Zarnowitz (2003). Within the factor based approaches the possibility of measurement error in the components of the leading index, due e.g. to data revisions, can be also formally taken into account, as discussed in Section 2.1, but in practice also the resulting composite indexes require later revisions. Yet, both for the traditional and for the more sophisticated methods, the revisions in the composite indexes due to the use of later releases of their components are minor.

The final property vi), a smooth evolution in the leading indicator, can require a careful choice of variable transformations and/or filter. In particular, the filtering procedures discussed in Section 2.2 to enhance the business cycle characteristics of the coincident indicator can be applied also to the leading indicators, and in general should be if the target variable is filtered. In general, they can provide improvements with respect to the standard choice of month to month differences of the leading indicator. Also, longer differences can be useful to capture sustained growth or lack of it, see e.g. Birchenhall et al. (1999), or differences with respect to the previous peak or trough to take into consideration the possible non-stationary variations of values at turning points, see e.g. Chin et al. (2000).

3.2 Composite indexes

The use of a single leading indicator is dangerous because economic theory and experience teach that recessions can have different sources and characteristics. For example, the twin US recessions of the early 80's were mostly due to tight monetary policy, that of 1991 to a deterioration in the expectations climate because of the first Iraq war, and that of 2001 to the bursting of the stock market bubble and, more generally, to over-investment, see e.g. Stock and Watson (2003b). In the euro area, the three latest recessions according to the CEPR dating are also rather different, with the one in 1974 lasting only three quarters and characterized by synchronization across countries and coincident variables, as in 1992-93 but contrary to the longer recession that started at the beginning of 1980 and lasted 11 quarters.

A combination of leading indicators into composite indexes can therefore be more useful in capturing the signals coming from different sectors of the economy. The construction of a composite index requires several steps and can be undertaken either in a non model based framework or with reference to a specific econometric model of the evolution of the leading indicators, possibly jointly with the target variable. The latter case is analyzed in details in the next Section, while here we focus on the non model based approach to composite leading indexes construction, see e.g. Niemira and Klein (1994, Ch.3) for details. Examples are provided in the next subsection.

The first element is the selection of the index components. Each component should satisfy the criteria mentioned in the previous subsection. In addition, a balanced representation of all the sectors of the economy should be achieved, or at least of those more closely related to the target variable.

The second element is the transformation of the index components. This includes the transformations discussed above, namely, seasonal adjustment, outlier removal, treatment of measurement error in first releases of indicators subject to subsequent revision, and possibly forecast of unavailable most recent observations for some indicators. These steps can be implemented either in a univariate framework, mostly by exploiting univariate time series models for each indicator, or in a multivariate context. In addition, the transformed indicators should be made comparable to be included in a single index. Therefore, they are typically detrended (using different procedures such as differencing, regression on deterministic trends, or the application of more general band-pass filters), possibly smoothed to eliminate high frequency movements (using moving averages or, again, band pass filters), and standardized to make their amplitudes similar or equal (as in the case of the composite coincident index of Section 2.1).

The final element for the construction of a composite leading index is the choice of a weighting scheme. The typical choice, once the components have been standardized, is to give them equal weights. This seems a sensible averaging scheme in this context, unless there are particular reasons to give larger weights to specific sectors, depending on the target variable or on additional information on the economic situation.

From an econometric point of view, composite indexes constructed following the procedure sketched above are subject to several criticisms, some of which are derived in a formal framework in Emerson and Hendry (1996). First, even though the single indicators are typically chosen according to some formal or informal bivariate analysis of their relationship with the target variable, there is no explicit reference to the target variable in the construction of the composite index, e.g. in the choice of the weighting scheme. Second, the weighting scheme is fixed over time, with periodic revisions mostly due either to data issues, such as changes in the production process of an indicator, or to the past unsatisfactory performance of the index. Endogenously changing weights that track the possibly varying relevance of the single indicators over the business cycle and in the presence of particular types of shocks could produce better results, even though their derivation is difficult. Third, lagged values of the target variable are typically not included in the leading index, while there can be economic and statistical reasons underlying the persistence of the target variable that would favor such an inclusion. Fourth, lagged values of the single indicators are typically not used in the index, while they could provide relevant information, e.g. because not only the point value of an indicator but also its evolution over a period of time matter for anticipating the future behavior of the target variable. Fifth, if some indicators and the target variable are cointegrated, the presence of short run deviations from the long run equilibrium could provide useful information on future movements of the target variable. Finally, since the index is a forecast for the target variable, standard errors should be also provided, but their derivation is virtually impossible in the non model based context because of the lack of a formal relationship between the index and the target.

The main counterpart of these problems is simplicity. Non model based indexes are easy to build, easy to explain, and easy to interpret, which are very valuable assets, in particular for the general public and for policy makers. Moreover, simplicity is often a plus also for forecasting. With this method there is no estimation uncertainty, no major problems of overfitting, and the literature on forecast pooling suggests that equal weights work pretty well in practice, see e.g. Stock and Watson (2003a), even though here variables rather than forecasts are pooled.

Most of the issues raised for the non model based composite indexes are addressed by the model based procedures described in the next Section, which in turn are in general much more complicated and harder to understand for the general public. Therefore, while from the point of view of academic research and scientific background of the methods there is little to choose, practitioners may well decide to base their preferences on the practical forecasting performance of the two approaches to composite index construction.

3.3 Examples

To provide an illustration of the points made in this Section, we now analyze the indicator selection process for Stock and Watson's (1989, SW) model based composite leading index, described in details later on in Section 4.2, and the construction of two non model based indexes for the US produced by official agencies, the Conference Board, CLI_{CB}, and the OECD, CLI_{OECD}.

SW started with a rather large dataset of about 280 series, yet smaller than Mitchell and Burns' original selection of 487 candidate indicators. The series can be divided into ten groups: "measures of output and capacity utilization; consumption and sales; inventories and orders; money and credit quantity variables; interest rates and asset prices; exchange rates and foreign trade; employment, earnings and measures of the labor force; wages and prices; measures of government fiscal activity; and other variables", SW (p.365).

The bivariate relationships between each indicator, properly transformed, and the growth of the CCI_{DOC} were evaluated using frequency domain techniques (the coherence and the phase lead), and time domain techniques (Granger causality tests and marginal predictive content for CCI_{DOC} beyond that of CLI_{DOC}). The choice of CCI_{DOC} rather than CCI_{SW} as the target variable can raise some doubts, but the latter was likely not developed yet at the time, and in addition the two composite coincident indexes are highly correlated. Some series were retained even if they performed poorly on the basis of the three criteria listed above, because either economic theory strongly supported their inclusion or they were part of the CLI_{DOC}. After this first screening, 55 variables remained in the list of candidate components of the composite leading index.

It is interesting that SW mentioned the possibility of using all the 55 series for the construction of an index, but abandoned the project for technical reasons (at the time construction of a time series model for all these variables was quite complicated) and because it would be difficult to evaluate the contribution of each component to the index. About ten years later, the methodology to address the former issue became available, see Stock and Watson (2001a, 2001b) and the discussion in Section 4.2 below, but the latter issue remains, the trade-off between parsimony and broad coverage of the index is still unresolved.

The second indicator selection phase is based on a step-wise regression procedure. The dependent variable is $CCI_{SWt+6} - CCI_{SWt}$ i.e., the six months growth rate in the SW composite coincident index, that is also the target variable for SW composite leading index, see Section 4.2. Different sets of variables (including their lags as selected by the AIC) are used as regressors, variables in each set are retained on the basis of their marginal explanatory power, the best variables in each original set are grouped into other sets of regressors, and the procedure is repeated until a small number of indicators remains in the list.

At the end, seven variables (and their lags) were included in the composite index, as listed in Table 1 in SW. They are: i) an index of new private housing authorized, ii) the growth rate of manufacturers' unfilled orders for durable goods industries, iii) the growth rate in a trade weighted nominal exchange rate, iv) the growth rate of part-time work in non-agricultural industries, v) the difference of the yield on constant-maturity portfolio of 10-years US treasury bonds, vi) the spread between interest rates on 6-months corporate paper and 6-months US treasury bills, vii) the spread between the yield on 10-years and 1-year US Treasury bonds. The only change in the list so far took place in 1997, when the maturity in vi) became 3 months. SW also discussed theoretical explanations for the inclusion of these variables (and exclusion of others). As mentioned in the first subsection, the most innovative variables in SW's CLI_{SW} are the financial spreads, whose forecasting ability became the focus of theoretical and empirical research in subsequent years. Yet, following an analysis of the performance of their CLI_{SW} during the 1990 recession, see Section 5.2.3, Stock and Watson (1992) also introduced a non-financial based index ($CLI2_{SW}$).

A potential problem of the extensive variable search underlying the final selection of index components, combined with parameter estimation, is overfitting. Yet, when SW checked the overall performance of their selection procedure using Monte Carlo simulations, the results were satisfactory. Even better results were obtained by Hendry and Krolzig (1999, 2001) for their automated model selection procedure, PcGets, see Banerjee and Marcellino (2003) for an application to leading indicator selection for the US.

A final point worth noting about SW's indicator selection procedure is the use of variable transformations. First, seasonally adjusted series are used. Second, a stationarity transformation is applied for the indicator to have similar properties as the target. Third, some series are smoothed because of high frequency noise, in particular, ii), iii), iv), and v) in the list above. The adopted filter is $f(L) = 1 + 2L + 2L^2 + L^3$. Such a filter is chosen with reference to the target variable, the 6-month growth of CCI, and to the use of first differenced indicators, since $f(L)(1-L)$ is a band-pass filter with gains concentrated at periods of four months to one year. Finally, if the most recent values of some of the seven indicators are not available, they are substituted with forecasts in order to be able to use as timely information as possible. Zarnowitz and Braun (1989), in their comment to SW, pointed out that smoothing the indicators contributes substantially to the good forecasting performance of SW's CLI, combined with the use of the most up-to-date information.

The practice of using forecasts when timely data are not available is now supported also for the CLI_{CB}, see McGuckin et al. (2003), but not yet implemented in the published version of the index. The latter is computed following the same steps as for the coincident index, the CCI_{CB} described in Section 2.1, but with a different choice of components. In particular, the single indicators combined in the index include average weekly hours, manufacturing; average weekly initial claims for unemployment insurance; manufacturers' new orders, consumer good and materials (in 1996$); vendor performance, slower deliveries diffusion index; manufacturers' new orders, non-defense capital goods; building permits, new private housing units; stock prices, 500 common stocks; money supply (in 1996$); interest rate spread, 10-year Treasury bond less federal funds; and the University of Michigan's index of consumer expectations.

This list originates from the original selection of Mitchell and Burns (1938), but only two variables passed the test of time: average weekly hours in the manufacturing sector and the Standard and Poor's stock index (that replaces the Dow Jones index of industrial common stock prices), see Moore (1983) for an historical perspective. Both variables are not included in the CLI_{SW}, since their marginal contribution in forecasting the 6-month growth of the CCI_{SW} is not statistically significant. Other major differences in the components of the two composite leading indexes are the inclusion in CLI_{CB} of M2 and of the index of consumer expectations (the relationship of M2 with the CCI_{SW} is found to be unstable, while consumer expectations were added to CLI_{CB} in the '90s so that the sample is too short for a significant evaluation of their role); and the exclusion from CLI_{CB} of an exchange rate measure and of the growth in part time work (yet, the former has a small weight in the CLI_{SW}, while the latter is well proxied by the average weekly hours in manufacturing and the new claims for unemployment insurance).

The final example we consider is the OECD composite short leading index, CLI_{OECD}, for the US (see www.oecd.org). Several points are worth making. First, the target is represented by the turning points in the growth cycle of industrial production, where the trend component is estimated using a modified version of the phase average trend (PAT) method developed at the NBER (see OECD (1987), Niemira and Klein (1994) for details), and the Bry-Boschan

(1971) methodology is adopted for dating peaks and troughs. All of these choices are rather questionable, since industrial production is a lower and lower share of GDP, though still one of the most volatile components, theoretically sounder filters such as those discussed in Section 2.2 are available for detrending, and more sophisticated procedures are available for dating, see again Section 2.2. On the other hand, since the OECD computes the leading index for a wide variety of countries, simplicity and robustness are also relevant for them.

Second, the criteria for the selection of the components of the index are broadly in line with those listed in Section 3.1. The seven chosen indicators as listed in the OECD web site include dwellings started; net new orders for durable goods, share price index; consumer sentiment indicator; weekly hours of work, manufacturing; purchasing managers index; and the spread of interest rates. Overall, there is a strong similarity with the elements of the CLI_{CB}.

Third, as for CLI_{CB}, the components are first standardized and then aggregated with equal weights. More precisely, each indicator is detrended with the PAT method; smoothed according to its months for cyclical dominance (MCD) values to reduce irregularity (see OECD (1987) for details); transformed to homogenize the cyclical amplitudes; standardized by subtracting the mean from the observed values and then dividing the resulting difference by the mean of the absolute values of the differences from the mean; and finally aggregated. When timely data for an indicator are not available, it is not included in the preliminary release of the composite leading index.

Finally, the composite index is adjusted to ensure that its cyclical amplitude on average agrees with that of the detrended reference series. The trend restored version of the index is also computed and published, to get comparability with the IP series.

To provide an empirical illustration of the issues discussed in this Section, in Figure 4 we graph four composite leading indexes for the US: the Conference Board's leading index (CLI_{CB}), the OECD leading index (CLI_{OECD}), the Economic Cycle Research Institute's (ECRI, see www.businesscycle.com) weekly leading index (CLI_{ECRI}), and a transformation of Stock and Watson's (1989) composite leading index ($TCLI_{SW}$), their leading index plus their coincident index that yields a 6-month ahead forecast for the level of the coincident index, see Section 3.2. All indexes are normalized to have zero mean and unit standard deviation. In the same figure we graph the NBER dated recessions (shaded areas).

Visual inspection suggests that the four indices move closely together, and their peaks anticipate NBER recessions. These issues are more formally evaluated in Tables 2 and 3. In Table 2 we report the correlations of the 6-month percentage changes of the four indices, which are indeed high, in particular when the '60s are excluded from the sample, the lowest value being 0.595 for CLI_{SW} and CLI_{ECRI}.

In Table 3 we present a descriptive analysis of the peak and trough structure of the four leading indexes (obtained with the AMP algorithm), compared either with the NBER dating or with the dating of the CCI_{CB} resulting from the AMP algorithm. The $TCLI_{SW}$ has the worst performance in terms of missed peaks and troughs, but it is worth recalling that the goal of the CLI_{SW} is not predicting turning points but the 6-month growth rate of the CCI_{SW}. The other three leading indexes missed no peaks or troughs, with the exception of the 2002 peak identified only by the AMP dating algorithm. Yet, they gave three false alarms, in 1966, 1984-85, and 1994-95. The average lead for recessions is about 9-10 months for all indexes (slightly shorter for $TCLI_{SW}$), but for expansions it drops to only 3-4 months for CLI_{OECD} and CLI_{ECRI}. Based on this descriptive analysis, the CLI_{CB} appears to yield the best overall leading performance. Yet, these results should be interpreted with care since they are obtained with the final release of the leading indicators rather than with real time data, see Section 5.

In Figure 5 we report the HP band pass filtered versions of the four composite leading indexes, with the AMP deviation cycle dating (shaded areas). Again the series move closely together, slightly less so for the HPBP-$TCLI_{SW}$, and their peaks anticipate dated recessions.

From Table 4, the HPBP-$TCLI_{SW}$ is the least correlated with the other indexes, correlation coefficients are in the range $0.60 - 0.70$, while for the other three indexes the lowest correlation is 0.882.

From Table 5, the ranking of the indexes in terms of lead-time for peaks and troughs is similar to that in Table 3. In this case there is no official dating of the deviation cycle, so that we use the AMP algorithm applied to the HPBP-CCI_{CB} as a reference. The HPBP-CLI_{CB} confirms its good performance, with an average lead time of 7 months for recessions, 10 months for expansions, and just one missed signal and two false alarms. The HPBP-CLI_{ECRI} is a closed second, while the HPBP-$TCLI_{SW}$ remains the worst, with 3-4 missed signals.

Finally, the overall good performance of the simple non model based CLI_{CB} deserves further attention. We mentioned that it is obtained by cumulating, using the formula in (1), an equal weighted average of the one month symmetric percent changes of ten indicators. The weighted average happens to have a correlation of 0.960 with the first principal component of the ten members of the CLI_{CB}. The latter provides a non parametric estimator for the factor in a dynamic factor model, see Section 4.2 and Stock and Watson (2001a, 2001b) for details. Therefore, the CLI_{CB} can be also considered as a good proxy for a factor model based composite leading indicator.

4 Prediction with leading indicators

Leading indicators are hardly of any use without a rule to transform them into a forecast for the target variable. These rules range from simple non-parametric procedures that monitor the evolution of the leading indicator and transform it into a recession signal, e.g. the three-consecutive-declines in the CLI_{CB} rule (e.g. Vaccara and Zarnowitz (1977)), to sophisticated non-linear models for the joint evolution of the leading indicators and the target variable, which can be used to predict growth rates, turning points, and expected duration of a certain business cycle phase. In this section we provide an overview of these methods. In particular, Section 4.1 deals with linear models, 4.2 with factor based models, 4.3 with Markov switching models, 4.4 with smooth transition models, 4.5 with neural network and non-parametric methods, 4.6 with binary models, and 4.7 with forecast pooling procedures. Examples are provided in the next Section, after having defined evaluation criteria for leading indicator based forecasts.

4.1 Linear models

A linear VAR provides the simplest model based framework to understand the relationship between coincident and leading indicators, the construction of regression based composite indexes, the role of the latter in forecasting, and the consequences of invalid restrictions or unaccounted cointegration.

Let us group the m coincident indicators in the vector x_t, and the n leading indicators in y_t. For the moment, (x_t, y_t) is weakly stationary and its evolution is described by the VAR(1):

$$\begin{aligned} \begin{pmatrix} x_t \\ y_t \end{pmatrix} &= \begin{pmatrix} c_x \\ c_y \end{pmatrix} + \begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} x_{t-1} \\ y_{t-1} \end{pmatrix} + \begin{pmatrix} e_{xt} \\ e_{yt} \end{pmatrix}, \\ \begin{pmatrix} e_{xt} \\ e_{yt} \end{pmatrix} &\sim i.i.d. \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{pmatrix} \right). \end{aligned} \tag{24}$$

It immediately follows that the expected value of x_{t+1} conditional on the past is

$$E(x_{t+1}|x_t, x_{t-1}, ...y_t, y_{t-1}, ...) = c_x + Ax_t + By_t, \tag{25}$$

so that for y to be a useful set of leading indicators it must be $B \neq 0$. When $A \neq 0$, lagged values of the coincident variables also contain useful information for forecasting. Both hypotheses are easily testable and, in case both $A = 0$ and $B = 0$ are rejected, a composite regression based leading indicator for x_{t+1} (considered as a vector) can be constructed as

$$CLI1_t = \widehat{c}_x + \widehat{A}x_t + \widehat{B}y_t, \tag{26}$$

where the $\widehat{}$ indicates the OLS estimator. Standard errors around this CLI can be constructed using standard methods for VAR forecasts, see e.g. Lütkepohl (2004). Moreover, recursive estimation of the model provides a convenient tool for continuous updating of the weights.

A similar procedure can be followed when the target variable is dated $t + h$ rather than t. For example, when $h = 2$,

$$\begin{aligned} CLI1_t^{h=2} &= \widehat{c}_x + \widehat{A}\widehat{x}_{t+1|t} + \widehat{B}\widehat{y}_{t+1|t} \\ &= \widehat{c}_x + \widehat{A}(\widehat{c}_x + \widehat{A}x_t + \widehat{B}y_t) + \widehat{B}(\widehat{c}_y + \widehat{C}x_t + \widehat{D}y_t). \end{aligned} \tag{27}$$

As an alternative, the model in (24) can be re-written as

$$\begin{pmatrix} x_t \\ y_t \end{pmatrix} = \begin{pmatrix} \widetilde{c}_x \\ \widetilde{c}_y \end{pmatrix} + \begin{pmatrix} \widetilde{A} & \widetilde{B} \\ \widetilde{C} & \widetilde{D} \end{pmatrix} \begin{pmatrix} x_{t-h} \\ y_{t-h} \end{pmatrix} + \begin{pmatrix} \widetilde{e}_{xt} \\ \widetilde{e}_{yt} \end{pmatrix} \tag{28}$$

where a $\widetilde{}$ indicates that the new parameters are a combination of those in (24), and $\widetilde{e}_{xt}$ and $\widetilde{e}_{yt}$ are correlated of order $h - 1$. Specifically,

$$\begin{aligned} \begin{pmatrix} \widetilde{c}_x \\ \widetilde{c}_y \end{pmatrix} &= \left(I + \begin{pmatrix} A & B \\ C & D \end{pmatrix} + ... + \begin{pmatrix} A & B \\ C & D \end{pmatrix}^{h-1} \right) \begin{pmatrix} c_x \\ c_y \end{pmatrix}, \\ \begin{pmatrix} \widetilde{A} & \widetilde{B} \\ \widetilde{C} & \widetilde{D} \end{pmatrix} &= \begin{pmatrix} A & B \\ C & D \end{pmatrix}^h, \\ \begin{pmatrix} \widetilde{e}_{xt} \\ \widetilde{e}_{yt} \end{pmatrix} &= \left(I + \begin{pmatrix} A & B \\ C & D \end{pmatrix} + ... + \begin{pmatrix} A & B \\ C & D \end{pmatrix}^{h-1} \right) \begin{pmatrix} e_{xt} \\ e_{yt} \end{pmatrix}. \end{aligned} \tag{29}$$

The specification in (28) can be estimated by OLS, and the resulting CLI written as

$$\widetilde{CLI1}_t^h = \widehat{\widetilde{c}}_x + \widehat{\widetilde{A}}x_t + \widehat{\widetilde{B}}y_t. \tag{30}$$

The main disadvantage of this latter method, often called dynamic estimation, is that a different model has to be specified for each forecast horizon h. On the other hand, no model is required for the leading indicators, and the estimators of the parameters in (28) can be more robust than those in (24) in the presence of mis-specification, see e.g. Clements and Hendry (1996) for a theoretical discussion and Marcellino, Stock and Watson (2004) for an extensive empirical analysis of the two competing methods (showing that dynamic estimation is on average slightly worse than the iterated method for forecasting US macroeconomic time series). For the sake of simplicity, in the rest of the paper we will focus on $h = 1$ whenever possible.

Consider now the case where the target variable is a composite coincident indicator,

$$CCI_t = wx_t, \tag{31}$$

where w is a $1 \times m$ vector of weights as in Section 3.2. To construct a model based CLI for the CCI in (31) two routes are available. First, and more common, we could model CCI_t and y_t with a finite order VAR, say

$$\begin{pmatrix} CCI_t \\ y_t \end{pmatrix} = \begin{pmatrix} d_{CCI} \\ d_y \end{pmatrix} + \begin{pmatrix} e(L) & F(L) \\ g(L) & H(L) \end{pmatrix} \begin{pmatrix} CCI_{t-1} \\ y_{t-1} \end{pmatrix} + \begin{pmatrix} u_{CCIt} \\ u_{yt} \end{pmatrix}, \tag{32}$$

where L is the lag operator and the error process is white noise. Repeating the previous procedure, the composite leading index for $h = 1$ is

$$CLI2_t = \widehat{d}_{CCI} + \widehat{e}(L)CCI_t + \widehat{F}(L)y_t. \tag{33}$$

Yet, in this case the VAR is only an approximation for the generating mechanism of (wx_t, y_t), since in general the latter should have an infinite number of lags or an MA component.

The alternative route is to stick to the model in (24), and construct the CLI as

$$CLI3_t = wCLI1_t, \tag{34}$$

namely, aggregate the composite leading indicators for each of the components of the CCI, using the same weights as in the CCI. Lütkepohl (1987) showed in a related context that in general aggregating the forecasts ($CLI3$) is preferable than forecasting the aggregate ($CLI2$) when the variables are generated by the model in (24), while this is not necessarily the case if the model in (24) is also an approximation and/or the x variables are subject to measurement error. Stock and Watson (1992) overall found little difference in the performance of $CLI2$ and $CLI3$.

Both $CLI2$ and $CLI3$ are directly linked to the target variable, incorporate distributed lags of both the coincident and the leading variables (depending on the lag length of the VAR), the weights can be easily periodically updated using recursive estimation of the model, and standard errors around the point forecasts (or the whole distribution under a distributional assumption for the error process in the VAR) are readily available. Therefore, this simple linear model based procedure already addresses several of the main criticisms to the non model based composite index construction, see Section 3.2.

In this context the dangers of using a simple average of the y variables as a composite leading index are also immediately evident, since the resulting index can provide an inefficient forecast of the CCI unless specific restrictions on the VAR coefficients in (24) are satisfied. In particular, indicating by i_n a $1 \times n$ vector with elements equal to $1/n$, the equal weight composite leading index

$$CLI_{EWt} = i_n y_t \tag{35}$$

is optimal and coincides with $CLI3$ if and only if

$$wc_x = 0, \quad wA = 0, \quad wB = i_n, \tag{36}$$

which imposes $1 + m + n$ restrictions on the parameters of the x equations in (24). In higher order VARs, the product of the weights w and the coefficients of longer lags of x and y in the x equations should also be equal to zero. Notice that these are all testable assumptions as long as $m + n$ is small enough with respect to the sample size to leave sufficient degrees of freedom for the VAR parameter estimation. For example, in the case of the Conference Board, $m+n = 14$ and monthly data are available for about 45 years for a total of more than 500 observations. Auerbach (1982) found that a regression based CLI in sample performed better than the the equal weighted CLI_{CB} for industrial production and the unemployment rate, but not out of sample.

If the restrictions in (36) are not satisfied but it is desired to use in any case CLI_{EW} (or more generally a given CLI) to forecast the CCI, it can be possible to improve upon its performance by constructing a VAR for the two composite indexes CCI and CLI_{EW} $(wx_t, i_n y_t)$, say

$$\begin{pmatrix} CCI_t \\ CLI_{EWt} \end{pmatrix} = \begin{pmatrix} f_{CCI} \\ f_{CLI_{EW}} \end{pmatrix} + \begin{pmatrix} e(L) & f(L) \\ g(L) & h(L) \end{pmatrix} \begin{pmatrix} CCI_{t-1} \\ CLI_{EWt-1} \end{pmatrix} + \begin{pmatrix} v_{CCIt} \\ v_{CLI_{EW}t} \end{pmatrix} \tag{37}$$

and construct the new composite index as

$$CLI4_t = \widehat{f}_{CCI} + \widehat{e}(L)CCI_t + \widehat{f}(L)CLI_{EWt}. \tag{38}$$

This is for example the methodology adopted by Kock and Rasche (1988), who analyzed a VAR for IP, as a coincident indicator, and the equal weighted DOC leading index. Since $CLI4$ has a dynamic structure and exploits also past information in the CCI, it can be expected to improve upon CLI_{EW}. Moreover, since the VAR in (37) is much more parsimonious than both (24) and (32), $CLI4$ could perform in practice even better than the other composite indexes, in particular in small samples.

A point that has not deserved attention in the literature but can be of importance is the specification of the equations for the (single or composite) leading indicators. Actually, in all the models we have considered so far, the leading variables depend on lags of the coincident ones, which can be an unreliable assumption from an economic point of view. For example, the interest rate spread depends on future expected short term-interest rates and the stock market index on future expected profits and dividends, and these expectations are positively and highly correlated with the future expected overall economic conditions. Therefore, the leading variables could depend on future expected coincident variables rather than on their lags. For example, the equations for y_t in the model for (x_t, y_t) in (24) could be better specified as:

$$y_t = c_y + Cx^e_{t+1|t-1} + Dy_{t-1} + e_{yt}, \tag{39}$$

where $x^e_{t+1|t-1}$ indicates the expectation of x_{t+1} conditional on information available in period $t-1$. Combining these equations with those for x_t in (24), it is possible to obtain a closed form expression for $x^e_{t+1|t-1}$, which is

$$x^e_{t+1|t-1} = (I - BC)^{-1}(c_x + Ac_x + Bc_y + A^2x_{t-1} + (AB + BD)y_{t-1}). \tag{40}$$

Therefore, a VAR specification such as that in (24) can be also considered as a reduced form of a more general model where the leading variables depend on expected future coincident variables. A related issue is whether the coincident variables, x_t, could also depend on their future expected values, as it often results in new-Keynesian models, see e.g. Walsh (2003). Yet, the empirical evidence in Fuhrer and Rudebusch (2002) provides little support for this hypothesis.

Another assumption we have maintained so far is that both the coincident and the leading variables are weakly stationary, while in practice it is likely that the behaviour of most of these variables is closer to that of integrated process. Following Sims, Stock and Watson (1990), this is not problematic for consistent estimation of the parameters of VARs in levels such as (24), and therefore for the construction of the related $CLIs$, even though inference is complicated and, for example, hypotheses on the parameters such as those in (36) could not be tested using standard asymptotic distributions. An additional complication is that in this literature, when the indicators are I(1), the VAR models are typically specified in first differences rather than in levels, without prior testing for cointegration. Continuing the VAR(1) example, the adopted model would be

$$\begin{pmatrix} \Delta x_t \\ \Delta y_t \end{pmatrix} = \begin{pmatrix} c_x \\ c_y \end{pmatrix} + \begin{pmatrix} e_{xt} \\ e_{yt} \end{pmatrix}, \tag{41}$$

rather than possibly

$$\begin{aligned} \begin{pmatrix} \Delta x_t \\ \Delta y_t \end{pmatrix} &= \begin{pmatrix} c_x \\ c_y \end{pmatrix} - \left(\begin{pmatrix} I_m & 0 \\ 0 & I_n \end{pmatrix} - \begin{pmatrix} A & B \\ C & D \end{pmatrix} \right) \begin{pmatrix} x_{t-1} \\ y_{t-1} \end{pmatrix} + \begin{pmatrix} e_{xt} \\ e_{yt} \end{pmatrix} \\ &= \begin{pmatrix} c_x \\ c_y \end{pmatrix} - \alpha\beta' \begin{pmatrix} x_{t-1} \\ y_{t-1} \end{pmatrix} + \begin{pmatrix} e_{xt} \\ e_{yt} \end{pmatrix}, \end{aligned} \tag{42}$$

where β is the matrix of cointegrating coefficients and α contains the loadings of the error correction terms. As usual, omission of relevant variables yields biased estimators of the parameters of the included regressors, which can translate into biased and inefficient composite leading indicators. See Emerson and Hendry (1996) for additional details and generalizations and, e.g., Clements and Hendry (1999) for the consequences of omitting cointegrating relations when forecasting. As long as $m+n$ is small enough with respect to the sample size, the number and composition of the cointegrating vectors can be readily tested, see e.g. Johansen (1988) for tests within the VAR framework, and the specification in (42) used as a basis to construct model based $CLIs$ that take also cointegration into proper account. Hamilton and Perez-Quiros (1996) found cointegration to be important for improving the forecasting performance of the CLI_{DOC}.

Up to now we have implicitly assumed, as it is common in most of the literature that analyzes $CCIs$ and $CLIs$ within linear models, that the goal of the composite leading index is forecasting a continuous variable, the CCI. Yet, leading indicators were originally developed for forecasting business cycle turning points. Simulation based methods can be used to derive forecasts of a binary recession/expansion indicator, and these in turn can be exploited to forecast the probability that a recession will take place within, or at, a certain horizon.

Let us consider the model in (32) and assume that the parameters are known and the errors are normally distributed. Then, drawing random numbers from the joint distribution of the errors for period $t+1, ..., t+n$ and solving the model forward, it is possible to get a set of simulated values for $(CCI_{t+1}, \Delta y_{t+1}), ..., (CCI_{t+n}, \Delta y_{t+n})$. Repeating the exercise many times, a histogram of the realizations provides an approximation for the conditional distribution of $(CCI_{t+1}, \Delta y_{t+1}), ..., (CCI_{t+n}, \Delta y_{t+n})$ given the past. Given this distribution and a rule to transform the continuous variable CCI into a binary recession indicator, e.g. the three months negative growth rule, the probability that a given future observation can be classified as a recession is computed as the fraction of the relevant simulated future values of the CCI that satisfy the rule.

A related problem that could be addressed within this framework is forecasting the beginning of the next recession, which is given by the time index of the first observation that falls into a recessionary pattern. Assuming that in period t the economy is in expansion, the probability of a recession after q periods, i.e., in $t+q$, is equal to the probability that $CCI_{t+1}, ..., CCI_{t+q-1}$ belong to an expansionary pattern while CCI_{t+q} to a recessionary one.

The procedure can be easily extended to allow for parameter uncertainty by drawing parameter values from the distribution of the estimators rather than treating them as fixed. Normality of the errors is also not strictly required since re-sampling can be used, see e.g. Wecker (1979), Kling (1987) and Fair (1993) for additional details and examples.

Bayesian techniques are also available for forecasting turning points in linear models, see e.g. Geweke and Whiteman (2004). In particular, Zellner and Hong (1991) and Zellner, Hong and Gulati (1990) addressed the problem in a decision-theoretic framework, using fixed parameter AR models with leading indicators as exogenous regressors. In our notation, the model can be written as

$$x_t = z_t^{'}\beta + u_t, \quad u_t \sim i.i.d.N(0, \sigma^2), \tag{43}$$

where $z_t^{'} = (x_{t-1}, y_{t-1})$, x_t is a univariate coincident variable or index, y_t is the $1 \times n$ vector of leading indicators, and β is a $k \times 1$ parameter vector, with $k = n + 1$.

Zellner et al. (1990, 1991) used annual data and declared a downturn (DT) in year $T+1$ if the annual growth rate observations satisfy

$$x_{T-2}, x_{T-1} < x_T > x_{T+1}, \tag{44}$$

while no downturn (NDT) happens if

$$x_{T-2}, x_{T-1} < x_T \leq x_{T+1}. \tag{45}$$

Similar definitions were proposed for upturns and no upturns.

The probability of a DT in $T+1$, p_{DT}, can be calculated as

$$p_{DT} = \int_{-\infty}^{x_T} p(x_{T+1}|A_1, D_T) dx_{T+1}, \tag{46}$$

where A_1 indicates the condition $(x_{T-2}, x_{T-1} < x_T)$, D_T denotes the past sample and prior information as of period T, and p is the predictive probability density function (pdf) defined as

$$p(x_{T+1}|D_T) = \int_\theta f(x_{T+1}|\theta, D_T)\pi(\theta|D_T)d\theta, \tag{47}$$

where $f(x_{T+1}|\theta, D_T)$ is the pdf for x_{T+1} given the parameter vector $\theta = (\beta, \sigma^2)$ and D_T, while $\pi(\theta|D_T)$ is the posterior pdf for θ obtained by Bayes' Theorem.

The predictive pdf is constructed as follows. First, natural conjugate prior distributions are assumed for β and σ, namely, $p(\beta|\sigma) \sim N(0, \sigma^2 I \times 10^6)$ and $p(\sigma) \sim IG(v_0 s_0)$, where IG stands for inverted gamma and v_0 and s_0 are very small numbers, see e.g. Canova (2004, Ch.9) for details. Second, at $t = 0$, the predictive pdf $p(x_1|D_0)$ is a Student-t, namely, $t_{v_0} = (x_1 - z_1^{'}\widehat{\beta}_0)/s_0 a_0$ has a univariate Student-t density with v_0 degrees of freedom, where $a_0^2 = 1 + z_1^{'} z_1 10^6$ and $\widehat{\beta}_0 = 0$. Third, the posterior pdfs obtained period by period using the Bayes' Theorem are used to compute the period by period predictive pdfs. In particular, the predictive pdf for x_{T+1} is again Student-t and

$$t_{v_T} = (x_{T+1} - z_{T+1}^{'}\widehat{\beta}_T)/s_T a_T \tag{48}$$

has a univariate Student-t pdf with v_T degrees of freedom, where

$$\begin{aligned}
\widehat{\beta}_T &= \widehat{\beta}_{T-1} + (Z_{T-1}^{'} Z_{T-1})^{-1} z_T (x_T - z_T^{'}\widehat{\beta}_{T-1})/[1 + z_T^{'}(Z_T^{'} Z_T)^{-1} z_T], \\
a_T^2 &= 1 + z_{T+1}^{'}(Z_T^{'} Z_T)^{-1} z_{T+1}, \\
v_T &= v_{T-1} + 1, \\
v_T s_T^2 &= v_{T-1} s_{T-1} + (x_T - z_T^{'}\widehat{\beta}_T)^2 + (\widehat{\beta}_T - \widehat{\beta}_{T-1})^{'} Z_{T-1}^{'} Z_{T-1} (\widehat{\beta}_T - \widehat{\beta}_{T-1}),
\end{aligned}$$

and $Z_T^{'} = (z_T, z_{T-1}, ..., z_1)$. Therefore, $\Pr(x_{T+1} < x_T|D_T) = \Pr(t_{v_T} < (x_T - z_{T+1}^{'}\widehat{\beta}_T)/s_T a_T|D_T)$, which can be analytically evaluated using the Student-t distribution with v_T degrees of freedom.

Finally, if the loss function is symmetric (i.e. the loss from wrongly predicting NDT in the case of DT is the same as predicting DT in the case of NDT), then a DT is predicted in period $T+1$ if $p_{DT} > 0.5$. Otherwise, the cut-off value depends on the loss structure, see also Section 4.6.

While the analysis in Zellner et al. (1990) is univariate, the theory for Bayesian VARs is also well developed, starting with Doan, Litterman and Sims (1984). A recent model in this class was developed by Zha (1998) for the Atlanta FED, and its performance in turning point forecasting is evaluated by Del Negro (2001). In this case the turning point probabilities are computed by simulations from the predictive pdf rather than analytically, in line with the procedure illustrated above in the classical context.

To conclude, a common problem of VAR models is their extensive parameterization, which prevents the analysis of large data sets. Canova and Ciccarelli (2001, 2003) proposed

Bayesian techniques that partly overcome this problem, extending previous analysis by e.g. Zellner, Hong and Min (1991) and providing applications to turning point forecasting, see Canova (2004, Ch.10) for an overview. As an alternative, factor models can be employed, as we discuss in the next subsection.

4.2 Factor based models

The idea underlying Stock and Watson's (1989, SW) methodology for the construction of a CCI, namely that a single common force drives the evolution of several variables, can be also exploited to construct a CLI. In particular, if the single leading indicators are also driven by the (leads of the) same common force, then a linear combination of their present and past values can contain useful information for predicting the CCI.

To formalize the intuition above, following SW, the equation (4) in Section 2.2 is substituted with

$$\Delta C_t = \delta_C + \lambda_{CC}(L)\Delta C_{t-1} + \Lambda_{Cy}(L)\Delta y_{t-1} + v_{ct}. \tag{49}$$

and, to close the model, equations for the leading indicators are also added

$$\Delta y_t = \delta_y + \lambda_{yC}(L)\Delta C_{t-1} + \Lambda_{yy}(L)\Delta y_{t-1} + v_{yt}, \tag{50}$$

where v_{ct} and v_{yt} are i.i.d. and uncorrelated with the errors in (3).

The model in (2), (3), (49), (50) can be cast into state space form and estimated by maximum likelihood through the Kalman filter. SW adopted a simpler two-step procedure, where in the first step the model (2), (3), (4) is estimated, and in the second step the parameters of (49), (50) are obtained conditional on those in the first step. This procedure is robust to mis-specification of the equations (49), (50), in particular the estimated CCI coincides with that in Section 2.2, but it can be inefficient when either the whole model is correctly specified or, at least, the lags of the leading variables contain helpful information for estimating the current status of the economy. Notice also that the "forecasting" system (49), (50) is very similar to that in (32), the main difference being that here C_t is unobservable and therefore substituted with the estimate obtained in the first step of the procedure, which is CCI_{SW}. Another minor difference is that SW constrained the polynomials $\lambda_{yC}(L)$ and $\Lambda_{yy}(L)$ to eliminate higher order lags, while $\lambda_{CC}(L)$ and $\Lambda_{Cy}(L)$ are left unrestricted, see SW for the details on the lag length determination.

The SW composite leading index is constructed as

$$CLI_{SW} = \widehat{C}_{t+6|t} - C_{t|t}, \tag{51}$$

namely, it is a forecast of the 6-momth growth rate in the CCI_{SW}, where the value in $t+6$ is forecasted and that in t is estimated. This is rather different from the NBER tradition, represented nowadays by the CLI_{CB} that, as mentioned, aims at leading turning points in the level of the CCI. Following the discussion in Section 2.1, focusing on growth rather than on levels can be more interesting in periods of prolonged expansions.

A few additional comments are in order about the SW's procedure. First, the leading indicators should depend on expected future values of the coincident index rather than on its lags, so that a better specification for (50) is along the lines of (39). Yet, we have seen that in the reduced form of (39) the leading indicators depend on their own lags and on those of the coincident variables, and a similar comment holds in this case. Second, the issue of parameter constancy is perhaps even more relevant in this enlarged model, and in particular for forecasting. Actually, in a subsequent (1997) revision of the procedure, SW made the deterministic component of (49), δ_C, time varying; in particular, it evolves according to a random walk. Third, dynamic estimation of equation (49) would avoid the need of (50). This would be particularly convenient in this framework where the dimension of y_t is rather large, and a single forecast horizon is considered, $h = 6$. Fourth, rather than directly forecasting the CCI_{SW}, the components of x_t could be forecasted and then aggregated into the composite index using the in sample weights, along the lines of (34). Fifth, while SW formally tested for lack of cointegration among the components of x_t, they did not do it among the elements of y_t, and of (x_t, y_t), namely, there could be omitted cointegrating relationships either among the leading indicators, or among them and the coincident indicators. Finally, the hypothesis of a single factor driving both the coincident and the leading indicators should be formally tested.

Otrok and Whiteman (1998) derived a Bayesian version of SW's CCI and CLI. As in the classical context, the main complication is the non-observability of the latent factor. To address this issue, a step-wise procedure is adopted where the posterior distribution of all unknown parameters of the model is determined conditional on the latent factor, then the conditional distribution of the latent factor conditional on the data and the other parameters is derived, the joint posterior distribution for the parameters and the factor is sampled using a Markov Chain Monte Carlo procedure using the conditional distributions in the first two steps, and a similar route is followed to obtain the marginal predictive pdf of the factor, which is used in the construction of the leading indicator, see Otrok and Whiteman (1998), Kim and Nelson (1998), Filardo and Gordon (1999) for details and Canova (2004, Ch.11) for an overview.

The SW's methodology could be also extended to exploit recent developments in the dynamic factor model literature. In particular, the step-wise procedure described in Section 2.1 to reduce the set of candidate leading indicators could be substituted with a factor model for all the indicators, and the estimated factors used to forecast the coincident index or its components. Let us sketch the steps of this approach, more details can be found in Stock and Watson (2004).

The model for the leading indicators in (50) can be replaced by

$$\Delta y_t = \Lambda f_t + \xi_t, \tag{52}$$

where the dimension of Δy_t can be very large, possibly larger than the number of observations (so that no sequential indicator selection procedure is needed), f_t is an $r \times 1$ vector of common factors (so that more than one factor can drive the indicators), and ξ_t is a vector containing the idiosyncratic component of each leading indicator. Precise moment conditions on f_t and ξ_t, and requirements on the loadings matrix Λ, are given in Stock and Watson (2002a, 2002b). Notice that f_t could contain contemporaneous and lagged values of factors, so that the model is truly dynamic even though the representation in (52) is static.

Though the model in (52) is a simple extension of that for the construction of the SW's composite coincident index in (2), its estimation is complicated by the possibly very large number of parameters, that makes maximum likelihood computationally not feasible. Therefore, Stock and Watson (2002a, 2002b) defined the factor estimators, $\widehat{f}_t$, as the minimizers of the objective function

$$V_{nT}(f,\Lambda)=\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}(y_{it}-\Lambda_i f_t)^2. \tag{53}$$

It turns out that the optimal estimators of the factors are the r eigenvectors corresponding to the r largest eigenvalues of the $T\times T$ matrix $n^{-1}\sum_{i=1}^{n}\underline{y}_i\underline{y}_i'$, where $\underline{y}_i=(y_{i1},...,y_{iT})$, and these estimators converge in probability to the space spanned by the true factors f_t. See Bai (2003) for additional inferential results, Bai and Ng (2002) for results related to the choice of the number of factors, r, Boivin and Ng (2003) for issues related to the choice of the size of the dataset (i.e. the number of leading indicators in our case), and Kapetanios and Marcellino (2003) for an alternative (parametric) estimation procedure.

The factors driving the leading indicators, possibly coinciding with (leads of) those driving the coincident indicators, can be related to the coincident composite index by replacing equation (49) with

$$\Delta C_t=\delta_C+\lambda_{CC}(L)\Delta C_{t-1}+\lambda_{Cy}(L)f_{t-1}+v_{ct}. \tag{54}$$

Another important result proved by Stock and Watson (2002a, 2002b) is that the factors in the equation above can be substituted by their estimated counterparts, $\widehat{f_t}$, without (asymptotically) modifying the mean square forecast error, see also Bai and Ng (2003) for additional results.

A forecasting procedure based on the use of (52) and (54), produced good results for the components of the CCI_{SW}, Stock and Watson (2002a, 2002b), but also for predicting macroeconomic variables for the Euro area, the UK, and the Accession countries, see, respectively, Marcellino, Stock and Watson (2003), Artis, Banerjee and Marcellino (2004), and Banerjee, Marcellino and Masten (2003). Yet, in these studies the set of indicators for factor extraction was not restricted to those with leading properties, and the target variable was not the composite coincident index. Camba-Mendez, Kapetanios, Smith and Weale (2001) used only leading indicators on the largest European countries for factor extraction (estimating iteratively the factor model cast in state-space form), and confirmed the good forecasting performance of the estimated factors when inserted in a VAR for predicting GDP growth.

The alternative factor based approach by FHLR described in Section 2.1 can be also used to construct a CLI. The leading variables are endogenously determined using the phase delay of their common components with respect to CCI_{FHLR} (the weighted average of the common components of interpolated monthly GDP for Euro area countries). An equal weight average of the resulting leading variables is the CLI_{FHLR}. Future values of the CCI_{FHLR} are predicted with a VAR for CCI_{FHLR}, CLI_{FHLR}. Further refinements of the methodology are presented in Forni et al. (2003a), with applications in Forni et al. (2003b).

All the factor based methods we have considered up to now focus on predicting continuous variables. Therefore, as in the case of linear models, we now discuss how to forecast discrete variables related to business cycle dynamics. In particular, we review the final important contribution of SW, further refined in Stock and Watson (1992), namely, the construction of a pattern recognition algorithm for the identification of recessions, and the related approach for computing recession probabilities.

As mentioned in Section 2.2, a recession is broadly defined by the three Ds: duration, a recession should be long enough; depth, there should be a substantial slowdown in economic activity; and diffusion, such a slowdown should be common to most sectors of the economy. Diffusion requires several series or a composite index to be monitored, and SW were in favor of the latter option, using their CCI (which, we recall, in the cumulated estimate of ΔC_t in equation (2)). Moreover, SW required a recession to be characterized by ΔC_t falling below a certain boundary value, b_{rt} (depth), for either (a) six consecutive months or (b) nine months with no more than one increase during the middle seven months (duration), where (b) is the same as requiring ΔC_t to follow for seven of nine consecutive months including the first and the last month. Expansions were treated symmetrically, with b_{et} being the counterpart of b_{rt}, and both b_{rt} and b_{et} were treated as i.i.d. normal random variables.

A particular month is classified as a recession if it falls in a recessionary pattern as defined above. In particular, suppose that it has to be decided whether month t belongs to a recessionary pattern. Because of the definition of a recessionary pattern, the longest span of time to be considered is given by $\Delta C_{t-8}, ..., \Delta C_{t-1}$ and $\Delta C_{t+1}, ..., \Delta C_{t+8}$. For example, it could be that ΔC_t is below the threshold b_{rt} and also $\Delta C_{t-i} < b_{rt-i}$ for $i = 1, ..., 5$; in this case the sequence $\Delta C_{t-5}, ..., \Delta C_t$ is sufficient to classify period t as a recession. But it could be that $\Delta C_{t-i} > b_{rt-i}$ for $i = 1, ..., 8$, $\Delta C_t < b_{rt}$, $\Delta C_{t+1} > b_{rt+1}$, and $\Delta C_{t+i} < b_{rt+i}$ for $i = 2, ..., 8$, which requires to consider the whole sequence of 17 periods $\Delta C_{t-8}, ...\Delta C_t, ..., \Delta C_{t+8}$ to correctly classify period t as a recession. Notice also that the sequence for ΔC_t has to be compared with the corresponding sequence of thresholds, $b_{rt-8}, ...b_{rt}, ..., b_{rt+8}$.

The binary recession indicator, R_t, takes the value 1 if ΔC_t belongs to a recessionary pattern, and 0 otherwise. The expansion indicator is defined symmetrically, but is also worth noting that the definition of recession is such that there can be observations that are classified neither as recessions nor as expansions. Also, there is no role for duration dependence or correlation, in the sense that the probability of recession is independent of the length of the current expansion or recession, and of past values of R_t.

The evaluation of the probability of recession in period $t + h$ conditional on information on the present and past of the CCI and of the leading indicators (and on the fact that $t + h$ belongs either to an expansionary or to a recessionary pattern), requires the integration of a 34-dimensional distribution, where 17 dimensions are due to the evaluation of an (estimated and forecasted) sequence for ΔC_t that spans 17 periods, and the remaining ones from integration with respect to the distribution of the threshold parameters. Stock and Watson (1992) described in details a simulation based procedure to perform numerically the integration, and reported results for their composite recession indicator, CRI_{SW}, that evaluates in real time the probability that the economy will be in a recession 6-months ahead.

Though a rule that transforms the CRI_{SW} into a binary variable is not defined, high values of the CRI_{SW} should be associated with realizations of recessions. Using the NBER dating as a benchmark, SW found the in-sample performance of the CRI quite satisfactory, as well as that of the CLI. Yet, out of sample, in the recessions of 1990 and 2001, both indicators failed to provide early strong warnings, an issue that is considered in more detail in Section 5.

To conclude, it is worth pointing out that the procedure underlying SW's CRI is not specific to their model. Given the definition of a recessionary pattern, any model that relates a CCI to a set of leading indicators or to a CLI can be used to compute the probability of recession in a given future period using the same simulation procedure as SW but drawing the random variables from the different model under analysis. The simplest case is when the model for the coincident indicator and the leading indexes is linear, which is the situation described at the end of the previous subsection.

4.3 Markov switching models

The MS model introduced in Section 2.1 to define an intrinsic coincident index, and in 2.2 to date the business cycle, can be also exploited to evaluate the forecasting properties of a single or composite leading indicator. In particular, a simplified version of the model proposed by Hamilton and Perez-Quiros (1996) can be written as

$$\begin{aligned} \Delta x_t - c_{s_t} &= a(\Delta x_{t-1} - c_{s_{t-1}}) + b(\Delta y_{t-1} - d_{s_{t+r-1}}) + u_{xt}, \\ \Delta y_t - d_{s_{t+r}} &= c(\Delta x_{t-1} - c_{s_{t-1}}) + d(\Delta y_{t-1} - d_{s_{t+r-1}}) + u_{yt}, \\ u_t &= (u_{xt}, u_{yt})' \sim i.i.d.N(0, \Sigma), \end{aligned} \tag{55}$$

where x and y are univariate, s_t evolves according to the constant transition probability Markov chain defined in (12), and the leading characteristics of y are represented not only by its influence on future values of x but also by its being driven by future values of the state variable, s_{t+r}.

The main difference between (55) and the MS model used in Section 2.1, equation (10), is the presence of lags and leads of the state variable. This requires to define a new state variable, s_t^*, such that

$$s_t^* = \begin{cases} 1 & \text{if } s_{t+r} = 1, s_{t+r-1} = 1, ..., s_{t-1} = 1, \\ 2 & \text{if } s_{t+r} = 0, s_{t+r-1} = 1, ..., s_{t-1} = 1, \\ 3 & \text{if } s_{t+r} = 1, s_{t+r-1} = 0, ..., s_{t-1} = 1, \\ \vdots & \qquad\qquad \vdots \\ 2^{r+2} & \text{if } s_{t+r} = 0, s_{t+r-1} = 0, ..., s_{t-1} = 0. \end{cases} \tag{56}$$

The transition probabilities of the Markov chain driving s_t^* can be derived from (12), and in the simplest case where $r = 1$ they are summarized by the matrix

$$P = \begin{pmatrix} p_{11} & 0 & 0 & 0 & p_{11} & 0 & 0 & 0 \\ p_{10} & 0 & 0 & 0 & p_{10} & 0 & 0 & 0 \\ 0 & p_{01} & 0 & 0 & 0 & p_{01} & 0 & 0 \\ 0 & p_{00} & 0 & 0 & 0 & p_{00} & 0 & 0 \\ 0 & 0 & p_{11} & 0 & 0 & 0 & p_{11} & 0 \\ 0 & 0 & p_{10} & 0 & 0 & 0 & p_{10} & 0 \\ 0 & 0 & 0 & p_{01} & 0 & 0 & 0 & p_{01} \\ 0 & 0 & 0 & p_{00} & 0 & 0 & 0 & p_{00} \end{pmatrix}, \tag{57}$$

whose i^{th}, j^{th} element corresponds to the probability that $s_t^* = i$ given that $s_{t-1}^* = j$.

The quantity of major interest is the probability that s_t^* assumes a certain value given the available information, namely,

$$\zeta_{t|t} = \begin{pmatrix} \Pr(s_t^* = 1|x_t, x_{t-1}, ..., x_1, y_t, y_{t-1}, ..., y_1) \\ \Pr(s_t^* = 2|x_t, x_{t-1}, ..., x_1, y_t, y_{t-1}, ..., y_1) \\ \vdots \\ \Pr(s_t^* = 2^{r+2}|x_t, x_{t-1}, ..., x_1, y_t, y_{t-1}, ..., y_1) \end{pmatrix}, \tag{58}$$

which is the counterpart of equation (13) in this more general context. The vector $\zeta_{t|t}$ and the conditional density of future values of the variables given the past, $f(x_{t+1}, y_{t+1}|\ s_{t+1}^*,$ $x_t, ..., x_1,\ y_t, ..., y_1)$, can be computed using the sequential procedure outlined in Section 2.1, see Hamilton and Perez-Quiros (1996), Krolzig (2003) for details. The latter can be used for forecasting future values of the coincident variable, the former to evaluate the current status of the economy or to forecast its future status up to period $t+r$. For example, the probability of being in a recession today is given by the sum of the rows of $\zeta_{t|t}$ corresponding to those values of s_t^* characterized by $s_t = 1$, while the probability of being in a recession in period $t + r$ is given by the sum of the rows of $\zeta_{t|t}$ corresponding to those values of s_t^* characterized by $s_{t+r} = 1$. To make inference on states beyond period $t+r$, it is possible to use the formula

$$\zeta_{t+m|t} = P^m \zeta_{t|t}, \tag{59}$$

which is a direct extension of the first row of (15).

Hamilton and Perez-Quiros (1996) found that their model provides only a weak signal of recession in 1960, 1970 and 1990. Moreover, the evidence in favor of the nonlinear cyclical factor is weak and the forecasting gains for predicting GNP growth or its turning point are minor with respect to a linear VAR specification. Even weaker evidence in favor of the MS specification was found when a cointegrating relationship between GNP and lagged CLI is included in the model. The unsatisfactory performance of the MS model could be due to the hypothesis of constant probability of recessions, as in the univariate context, see e.g., Filardo (1994). Evidence supporting this claim, based on the recession of 1990, is provided by Filardo and Gordon (1999).

Chauvet (1998) found a good performance also for the factor MS model in tracking the recession of 1990 using the proper version of $\zeta_{t|t}$ in that context. This is basically the only forecasting application of the factor MS models described in Section 2.1, so that further research is needed to close the gap. For example, SW's procedure for the CLI construction could be implemented using Kim and Nelson's (1998) MS version of the factor model, or a switching element could be introduced in the SW's VAR equations (49) and (50).

The MS model can be also used to derive analytic forecasts of recession (or expansion) duration. Suppose that x_t follows the simpler MS model in (10)-(12) and that it is known that in period t the economy is in a recession, i.e., $s_t = 1$. Then,

$$\Pr(s_{t+1} = 1|x_t, ..., x_1) = p_{11}, \tag{60}$$

$$\Pr(s_{t+2} = 1, s_{t+1} = 1|x_t, ..., x_1) = \Pr(s_{t+2} = 1|s_{t+1} = 1, x_t, ..., x_1) \Pr(s_{t+1} = 1|x_t, ..., x_1) = p_{11}^2,$$

$$\ldots$$

and the probability that the recession ends in period $t + n$ is

$$\Pr(s_{t+n} = 0, s_{t+n-1} = 1, ..., s_{t+1} = 1|x_t, ..., x_1) = (1 - p_{11})p_{11}^{n-1}. \tag{61}$$

If instead (12) is substituted with (19), i.e. the state probabilities are time-varying, then

$$\Pr(s_{t+n}=0,s_{t+n-1}=1,...,s_{t+1}=1|x_t,...,x_1)=(1-\widehat{p}_{11,t+n})\prod_{j=1}^{n-1}\widehat{p}_{11,t+j} \tag{62}$$

with

$$\widehat{p}_{11,t+j}=E\left(\frac{\exp(\theta y_{t+j-1})}{1+\exp(\theta y_{t+j-1})}\middle| x_t,...,x_1,y_t,...,y_1\right). \tag{63}$$

It follows that an estimator of the expected remaining duration of the recession, τ, in period t is given by

$$\widehat{\tau}=E(\tau|s_t=1)=\sum_{i=1}^{\infty}i(1-\widehat{p}_{11,t+i})\prod_{j=1}^{i-1}\widehat{p}_{11,t+j}, \tag{64}$$

which simplifies to

$$\widehat{\tau}=E(\tau|s_t=1)=\sum_{i=1}^{\infty}i(1-p_{11})p_{11}^{i-1}, \tag{65}$$

for constant probabilities. An interesting issue is therefore whether the leading indicators are useful to predict τ or not.

To conclude, Bayesian methods for the estimation of Markov switching models were developed by Albert and Chib (1993a), Mc Cullock and Tsay (1994), Filardo and Gordon (1994) and several other authors, see e.g. Filardo and Gordon (1999) for a comparison of bayesian linear, MS and factor models for coincident indicators, and Canova (2004, Ch.11) for an overview. Yet, to the best of our knowledge, there are no applications to forecasting turning points with Bayesian MS models while, for example, a bayesian replication of the Hamilton and Perez-Quiros (1996) exercise would be feasible and interesting.

4.4 Smooth transition models

In the class of MS models, the transition across states is abrupt and driven by an unobservable variable. As an alternative, in smooth transition (ST) models the parameters evolve over time at a certain speed, depending on the behavior of observable variables. In particular, the ST-VAR, that generalizes the linear model in (24) can be written as

$$\begin{aligned}\Delta x_t &= c_x + A\Delta x_{t-1} + B\Delta y_{t-1} + (c_x + A\Delta x_{t-1} + B\Delta y_{t-1})F_x + u_{xt},\\ \Delta y_t &= c_y + C\Delta x_{t-1} + D\Delta y_{t-1} + (c_y + C\Delta x_{t-1} + D\Delta y_{t-1})F_y + u_{yt},\\ u_t &= (u_{xt},u_{yt})' \sim i.i.d.N(0,\Sigma),\end{aligned} \tag{66}$$

where

$$F_x=\frac{\exp(\theta_0+\theta_1 z_{t-1})}{1+\exp(\theta_0+\theta_1 z_{t-1})},\quad F_y=\frac{\exp(\phi_0+\phi_1 z_{t-1})}{1+\exp(\phi_0+\phi_1 z_{t-1})}, \tag{67}$$

and z_{t-1} contains lags of x_t and y_t.

The smoothing parameters θ_1 and ϕ_1 regulate the shape of parameter change over time. When they are equal to zero, the model becomes linear, while for large values the model tends to a self-exciting threshold model (see e.g. Potter (1995), Artis, Galvao and Marcellino (2003)), whose parameters change abruptly as in the MS case. In this sense the ST-VAR provides a flexible tool for modelling parameter change.

The transition function F_x is related to the probability of recession. In particular, when the values of z_{t-1} are much smaller than the threshold value, θ_0, the value of F_x gets close to zero, while large values lead to values of F_x close to one. This is a convenient feature in particular when F_x depends on lags of y_t only, since it provides direct evidence on the usefulness of the leading indicators to predict recessions. As an alternative, simulation methods as in Section 4.1 can be used to compute the probabilities of recession.

Details on the estimation and testing procedures for ST models, and extensions to deal with more than two regimes or time-varying parameters, are reviewed, e.g., by van Dijk, Teräsvirta and Franses (2002), while Teräsvirta (2004) focuses on the use of ST models in forecasting. In particular, as it is common with nonlinear models, forecasting more than one-step ahead requires the use of simulation techniques, unless dynamic estimation is used as, e.g., in Stock and Watson (1999b) or Marcellino (2004b).

Univariate versions of the ST model using leading indicators as transition variables were analyzed by Granger, Teräsvirta and Anderson (1993), while Camacho (2004), Anderson and Vahid (2001) and Camacho and Perez-Quiros (2002) considered the VAR case. The latter authors found a significant change in the parameters only for the constant, in line with the MS specifications described in the previous subsection and with the time-varying constant introduced by SW to compute their CLI.

Finally, Bayesian techniques for the analysis of smooth transition models were developed by Lubrano (1995), and by Geweke and Terui (1993) and Chen and Lee (1995) for threshold models, see Canova (2004, Ch.11) for an overview. Yet, there are no applications to forecasting using leading indicators.

4.5 Neural networks and non-parametric methods

The evidence reported so far, and that in Section 5 below, is not sufficient to pin down the best parametric model to relate the leading to the coincident indicator, different sample periods or indicators can produce substantially different results. A possible remedy is to use artificial neural networks, which can provide a valid approximation to the generating mechanism of a vast class of non-linear processes, see e.g. Hornik, Stinchcombe and White (1989), and Swanson and White (1997), Stock and Watson (1999b), Marcellino (2004b) for their use as forecasting devices.

In particular, Stock and Watson (1999b) considered two types of univariate neural network specifications. The single layer model with n_1 hidden units (and a linear component) is

$$x_t = \beta_0' z_t + \sum_{i=1}^{n_1} \gamma_{1i} g(\beta_{1i}' z_t) + e_t, \tag{68}$$

where $g(z)$ is the logistic function, i.e., $g(z) = 1/(1 + e^{-z})$, and z_t includes lags of the dependent variable. Notice that when $n_1 = 1$ the model reduces to a linear specification with a logistic smooth transition in the constant. A more complex model is the double layer feedforward neural network with n_1 and n_2 hidden units:

$$x_t = \beta_0' z_t + \sum_{j=1}^{n_2} \gamma_{2j} g \left(\sum_{i=1}^{n_1} \beta_{2ji} g(\beta_{1i}' z_t) \right) + e_t. \tag{69}$$

The parameters of (68) and (69) can be estimated by non-linear least-squares, and forecasts obtained by dynamic estimation.

While the studies using NN mentioned so far considered point forecasts, Qi (2001) focused on turning point prediction. The model she adopted is a simplified version of (69), namely,

$$r_t = g \left(\sum_{i=1}^{n_1} \beta_{2i} g(\beta_{1i}' z_t) \right) + e_t, \tag{70}$$

where z_t includes lagged leading indicators in order to evaluate their forecasting role, and r_t is a binary recession indicator. Actually, since $g(.)$ is the logistic function, the predicted values from (70) are constrained to lie in the $[0, 1]$ interval. As for (68) and (69), the model is estimated by non-linear least-squares, and dynamic estimation is adopted when forecasting.

An alternative way to tackle the uncertainty about the functional form of the relationship between leading and coincident indicators is to adopt a non-parametric specification, with the cost for the additional flexibility being the required simplicity of the model. Based on the results from the parametric models they evaluated, Camacho and Perez-Quiros (2002) suggested the specification,

$$x_t = m(y_{t-1}) + e_t, \tag{71}$$

estimated by means of the Nadaraya-Watson estimator, see also Hardle and Vieu (1992). Therefore,

$$\widehat{x}_t = \left(\sum_{j=1}^{T} K\left(\frac{y_{t-1} - y_j}{h} \right) x_j \right) \Bigg/ \left(\sum_{j=1}^{T} K\left(\frac{y_{t-1} - y_j}{h} \right) \right), \tag{72}$$

where $K(.)$ is the Gaussian kernel and the bandwidth h is selected by leave-one-out cross validation.

The model is used to predict recessions according to the two negative quarters rule. For example,

$$\Pr(x_{t+2} < 0, x_{t+1} < 0 | y_t) = \int_{y_{t+2}<0} \int_{y_{t+1}<0} f(x_{t+2}, x_{t+1} | y_t) dx_{t+2} dx_{t+1}, \tag{73}$$

and the densities are estimated using an adaptive kernel estimator, see Camacho and Perez-Quiros (2002) for details.

Another approach that imposes minimal structure on the leading-coincident indicator connection is the pattern recognition algorithm proposed by Keilis-Borok, Stock, Soloviev and Mikhalev (2000). The underlying idea is to monitor a set of leading indicators, comparing their values to a set of thresholds, and when a large fraction of the indicators rise above the threshold a recession alarm, A_t, is sent. Formally, the model is

$$A_t = \begin{cases} 1 & \text{if } \sum_{k=1}^{N} \Psi_{kt} \geq N - b \\ 0 & \text{otherwise} \end{cases}, \tag{74}$$

where $\Psi_{kt} = 1$ if $y_{kt} \geq c_k$, and $\Psi_{kt} = 0$ otherwise. The salient features of this approach are the tight parameterization (only $N+1$ parameters, $b, c_1, ..., c_N$), which is in general a plus in forecasting, the transformation of the indicators into binary variables prior to their combination, (from y_{kt} to Ψ_{kt} and then summed with equal weights), and the focus on the direct prediction of recessions, A_t is a 0/1 variable.

Keilis-Borok et al. (2000) used 6 indicators: SW's CCI defined in Section 2.1 and five leading indicators, the interest rate spread, a short term interest rate, manufacturing and trade inventories, weekly initial claims for unemployment, and the index of help wanted advertising. They analyzed three different versions of the model in (74) where the parameters are either judgementally assigned or estimated by non-linear least squares, with or without linear filtering of the indicators, finding that all versions perform comparably and satisfactory, producing (in a pseudo out-of-sample context) an early warning of the five recessions over the period 1961 to 1990. Yet, the result should be interpret with care because of the use of the finally released data and of the selection of the indicators using full sample information, consider e.g. the use of the spread which was not common until the end of the '80s.

4.6 Binary models

In the models we have analyzed so far to relate coincident and leading indicators, the dependent variable is continuous, even though forecasts of business cycle turning points are feasible either directly (MS or ST models) or by means of simulation methods (linear or factor models). A simpler and more direct approach treats the business cycle phases as a binary variable, and models it using a logit or probit specification.

In particular, let us assume that the economy is in recession in period t, $R_t = 1$, if the unobservable variable s_t is larger than zero, where the evolution of s_t is governed by

$$s_t = \beta' y_{t-1} + e_t. \tag{75}$$

Therefore,

$$\Pr(R_t = 1) = \Pr(s_t > 0) = F(\beta' y_{t-1}), \tag{76}$$

where $F(.)$ is either the cumulative normal distribution function (probit model), or the logistic function (logit model). The model can be estimated by maximum likelihood, and the estimated parameters combined with current values of the leading indicators to provide an estimate of the recession probability in period $t+1$, i.e.,

$$\widehat{R}_{t+1} = \Pr(R_{t+1} = 1) = F(\widehat{\beta}' y_t). \tag{77}$$

The logit model was adopted, e.g., by Stock and Watson (1991) and the probit model by Estrella and Mishkin (1998), while Birchenhall et al. (1999) provided a statistical justification for the former in a Bayesian context (on the latter, see also Zellner and Rossi (1984) and Albert and Chib (1993b)). Binary models for European countries were investigated by Estrella and Mishkin (1997), Bernard and Gerlach (1998), Estrella, Rodrigues and Schich (2000), Birchenhall, Osborn and Sensier (2001), Osborn, Sensier and Simpson (2001), Moneta (2003).

Several points are worth discussing about the practical use of the probit or logit models for turning point prediction. First, often in practice the dating of R_t follows the NBER expansion/recession classification. Since there are substantial delays in the NBER's announcements, it is not known in period t whether the economy is in recession or not. Several solutions are available to overcome this problem. Either the model is estimated with data up to period $t-k$ and it is assumed that β remains constant in the remaining part of the sample; or R_t is substituted with an estimated value from an auxiliary binary model for the current status of the economy, e.g. using the coincident indicators as regressors, see e.g. Birchenhall et al. (1999); or one of the alternative methods for real-time dating of the cycle described in Section 2.2 is adopted.

Second, as in the case of dynamic estimation, a different model specification is required for each forecast horizon. For example, if a h-step ahead prediction is of interest, the model in (75) should be substituted with

$$s_t = \gamma_h' y_{t-h} + u_{t,h}. \tag{78}$$

This approach typically introduces serial correlation and heteroskedasticity into the error term $u_{t,h}$, so that the logit specification combined with nonlinear least squares estimation and robust estimation of the standard errors of the parameters can be preferred over standard maximum likelihood estimation, compare for example (70) in the previous subsection which can be considered as a generalization of (78). Notice also that $\widehat{\gamma}_h' y_{t-h}$ can be interpreted as a h-step ahead composite leading indicator. As an alternative, the model in (75) could be complemented with an auxiliary specification for y_t, say,

$$y_t = A y_{t-1} + v_t \tag{79}$$

so that

$$\Pr(R_{t+h}=1) = \Pr(s_{t+h}>0) = \Pr(\beta' A^{h-1} y_t + \eta_{t+h-1} + e_{t+h} > 0) = F_{\eta+e}(\beta' A^{h-1} y_t) \tag{80}$$

with $\eta_{t+h-1} = \beta' v_{t+h-1} + \beta' A v_{t+h-2} + ... + \beta' A^{h-1} v_t$. In general, the derivation of $F_{\eta+e}(.)$ is quite complicated, and the specification of the auxiliary model for y_t can introduce additional noise. Dueker (2003) extended and combined equations (75) and (79) into

$$\begin{pmatrix} s_t \\ y_t \end{pmatrix} = \begin{pmatrix} a & B \\ c & D \end{pmatrix} \begin{pmatrix} s_{t-1} \\ y_{t-1} \end{pmatrix} + \begin{pmatrix} e_{st} \\ e_{yt} \end{pmatrix}, \tag{81}$$

which is referred to as Qual-VAR because of its similarity with the models considered in Section 4.1. The model composed of the equation for s_t alone is the dynamic ordered probit studied by Eichengreen, Watson and Grossman (1985), who derived its likelihood and the related maximum likelihood estimators. Adding the set of equations for y_t has the main advantage of closing the model for forecasting purposes. Moreover, Dueker (2003) showed that the model can be rather easily estimated using Gibbs sampling techniques, and Dueker and Wesche (2001) found sizeable forecasting gains with respect to the standard probit model, in particular during recessionary periods.

Third, the construction of the probability of a recession within a certain period, say $t+2$, is complicated within the binary model framework. The required probability is given by $\Pr(R_{t+1}=0, R_{t+2}=1) + \Pr(R_{t+1}=1, R_{t+2}=0) + \Pr(R_{t+1}=1, R_{t+2}=1)$. Then, either from (78)

$$\Pr(R_{t+1}=1, R_{t+2}=1) = \Pr(s_{t+1}>0, s_{t+2}>0) = \Pr(u_{t+1,1} > -\gamma_1' y_t, u_{t+2,2} > -\gamma_2' y_t), \tag{82}$$

or from (80)

$$\Pr(R_{t+1}=1,R_{t+2}=1)=\Pr(s_{t+1}>0,s_{t+2}>0)=\Pr(\beta' y_t+e_{t+1}>0,\beta' Ay_t+\beta' v_{t+1}+e_{t+2}>0), \tag{83}$$

and similar formulae apply for $\Pr(R_{t+1}=0,R_{t+2}=1)$ and $\Pr(R_{t+1}=1,R_{t+2}=0)$. As long as the joint distributions in (82) and (83) are equivalent to the product of the marginal ones, as in this case assuming that v_t are uncorrelated with e_t, and the error terms are i.i.d., an analytic solution can be found. For higher values of h simulation methods are required. For example, a system made up of the models resulting using equation (78) for different values of h can be jointly estimated and used to simulate the probability values in (82). A similar approach can be used to compute the probability that an expansion (or a recession) will have a certain duration. A third, simpler alternative, is to define another binary variable directly linked to the event of interest, in this case,

$$R2_t=\begin{cases} 0 \text{ if no recession in period } t+1,\, t+2 \\ 1 \text{ if at least one recession in } t+1,\, t+2 \end{cases}, \tag{84}$$

and then model $R2_t$ with a probit or logit specification as a function of indicators dated up to period $t-1$. The problem of this approach is that it is not consistent with the model for R_t in equations (75), (76). The extent of the mis-specification should be evaluated in practice and weighted with the substantial simplification in the computations. A final, more promising, approach is simulation of the Qual-VAR model in (81), along the lines of the linear model in Section 4.1.

Fourth, an additional issue that deserves investigation is the stability of the parameters over time, and in particular across business cycle phases. Chin et al. (2000) proposed to estimate different parameters in expansions and recessions, using an exogenous classification of the states based on their definition of turning points. Dueker (1997, 2002) suggested to make the switching endogenous by making the parameters of (75) evolve according to a Markov chain. Both authors provided substantial evidence in favor of parameters instability.

Fifth, an alternative procedure to compute the probability of recession in period t consists of estimating logit or probit models for a set of coincident indicators, and then aggregating the resulting forecasts. The weights can be either those used to aggregate the indicators into a composite index, or they can be determined within a pooling context, as described in the next subsection.

Finally, as in the case of MS or ST models, the estimated probability of recession, $\widehat{r}_{t+1}$, should be transformed into a 0/1 variable using a proper rule. The common choices are of the type $\widehat{r}_t \geq c$ where c is either 0.5, a kind of uninformative Bayesian prior, or equal to the sample unconditional recession probability. Dueker (2002) suggested to make the cutoff values also regime dependent, say c_0 and c_1, and to compare the estimated probability with a weighted combination of c_0 and c_1 using the related regime probabilities. In general, as suggested e.g. by Zellner et al. (1990) and analyzed in details by Lieli (2004), the cutoff should be a function of the preferences of the forecasters.

4.7 Pooling

Since the pioneering work of Bates and Granger (1969), it is well known that pooling several forecasts can yield a mean square forecast error (msfe) lower than that of each of the individual forecasts, see Timmermann (2004) for a comprehensive overview. Hence, rather than selecting a preferred forecasting model, it can be convenient to combine all the available forecasts, or at least some subsets.

Several pooling procedures are available. The three most common methods in practice are linear combination, with weights related to the msfe of each forecast (see e.g. Granger and Ramanathan (1984)), median forecast selection, and predictive least squares, where a single model is chosen, but the selection is recursively updated at each forecasting round on the basis of the past forecasting performance.

Stock and Watson (1999b) and Marcellino (2004a) presented a detailed study of the relative performance of these pooling methods, using a large dataset of, respectively, US and Euro area macroeconomic variables, and taking as basic forecasts those produced by a range of linear and non-linear models. In general simple averaging with equal weights produces good results, more so for the US than for the Euro area. Stock and Watson (2003) focused on the role of pooling for GDP growth forecasts in the G-7 countries, using a larger variety of pooling methods, and dozens of models. They concluded that median and trimmed mean pooled forecasts produce a more stable forecasting performance than each of their component forecasts. Incidentally, they also found pooled forecasts to perform better than the factor based forecasts discussed in Section 4.2.

Camacho and Perez-Quiros (2002) focused on pooling leading indicator models, in particular they considered linear models, MS and ST models, probit specifications, and the non-parametric model described in Section 4.5, using regression based weights as suggested by Granger and Ramanathan (1984). Hence, the pooled forecast is obtained as

$$\widehat{x}_{t+1|t} = w_1\widehat{x}_{t+1|t,1} + w_2\widehat{x}_{t+1|t,2} + ... + w_p\widehat{x}_{t+1|t,p}, \tag{85}$$

and the weights, w_i, are obtained as the estimated coefficients from the linear regression

$$x_t = \omega_1\widehat{x}_{t|t-1,1} + \omega_2\widehat{x}_{t|t-1,2} + ... + \omega_p\widehat{x}_{t|t-1,p} + u_t \tag{86}$$

which is estimated over a training sample using the forecasts from the single models to be pooled, $\widehat{x}_{t|t-1,i}$, and the actual values of the target variable.

Camacho and Perez-Quiros (2002) evaluated the role of pooling not only for GDP growth forecasts but also for turning point prediction. The pooled recession probability is obtained as

$$\widehat{r}_{t+1|t} = F(a_1\widehat{r}_{t+1|t,1} + a_2\widehat{r}_{t+1|t,2} + ... + a_p\widehat{r}_{t+1|t,p}), \tag{87}$$

where $F(.)$ is the cumulative distribution function of a normal variable, and the weights, a_i, are obtained as the estimated parameters in the probit regression

$$r_t = F(\alpha_1\widehat{r}_{t|t-1,1} + \alpha_2\widehat{r}_{t|t-1,2} + ... + \alpha_p\widehat{r}_{t|t-1,p}) + e_t, \tag{88}$$

which is again estimated over a training sample using the recession probabilities from the single models to be pooled, $\widehat{r}_{t|t-1,i}$, and the actual values of the recession indicator, r_t.

The pooling method described above was studied from a theoretical point of view by Li and Dorfman (1996) in a Bayesian context. A more standard Bayesian approach to forecast combination is the use of the posterior odds of each model as weights, see e.g. Min and Zellner (1993). When all models have equal prior odds, this is equivalent to the use of the likelihood function value of each model as its weight in the pooled forecast.

5 Evaluation of leading indicators

In this section we deal with the evaluation of the forecasting performance of the leading indicators when used either in combination with simple rules to predict turning points, or as regressors in one of the models described in the previous Section to forecast either the growth rate of the target variable or its turning points. In the first subsection we consider methodological aspects. In the second subsection we discuss examples. In the final one we provide a rapid survey of the recent literature on the actual forecasting performance of single and composite leading indicators.

5.1 Methodology

A first assessment of the goodness of leading indicators can be based on standard in-sample specification and mis-specification tests of the models that relate the indicators to the target variable.

The linear model in (24) provides the simplest framework to illustrate the issues. A first concern is whether it is a proper statistical model of the relationships among the coincident and the leading variables. This requires the estimated residuals to mimic the assumed i.i.d. characteristics of the errors, the parameters to be stable over time, and the absence of non-linearity. Provided these hypotheses are not rejected, the model can be used to assess additional properties, such as Granger causality of the leading for the coincident indicators, or to evaluate the overall goodness of fit of the equations for the coincident variables (or for the composite coincident index). The model also offers a simple nesting framework to evaluate the relative merits of competing leading indicators, whose significance can be assessed by means of standard testing procedures. For a comprehensive analysis of the linear model see, e.g., Hendry (1995).

The three steps considered for the linear model, namely, evaluation of the goodness of the model from a statistical point of view, testing of hypotheses of interest on the parameters, and comparison with alternative specifications should be performed for each of the approaches listed in the previous Section. In particular, Hamilton and Raj (2002) and Raj (2002) provide up-to-date results for Markov-switching models, van Dijk, Teräsvirta and Franses (2002) for smooth transition models, while, e.g., Marcellino and Mizon (2004) present a general framework for model comparison.

Yet, in-sample analyses are more useful to highlight problems of a certain indicator or methodology than to provide empirical support in their favor, since they can be biased by over-fitting and related problems due to the use of the same data for model specification, estimation, and evaluation. A more sound appraisal of the leading indicators can be based on their out of sample performance, an additional reason for this being that forecasting is their main goal.

When the target is a continuos variables, such as the growth of a CCI over a certain period, standard forecast evaluation techniques can be used. In particular, the out-of-sample mean square forecast error (MSFE) or mean absolute error (MAE) provide standard summary measures of forecasting performance. Tests for equal forecast accuracy can be computed along the lines of Diebold and Mariano (1995), Clark and McCracken (2001), the standard errors around the MSFE of a model relative to a benchmark can be computed following West (1996), and tests for forecast encompassing can be constructed as in Clark and McCracken (2001). West (2004) provides an up-to-date survey of forecast evaluation techniques.

Moreover, as discussed in Section 4, simulation methods are often employed to compute the joint distribution of future values of the CCI to produce recession forecasts. Such a joint distribution can be evaluated using techniques developed in the density forecast literature, see e.g. Corradi and Swanson (2004).

When the target variable, R_t, is a binary indicator while the (out of sample) forecast is a probability of recession, P_t, similar techniques can be used since the forecast error is a continuous time variable. For example, Diebold and Rudebusch (1989) defined the accuracy of the forecast as

$$QPS = \frac{1}{T}\sum_{t=1}^{T} 2(P_t - R_t)^2, \tag{89}$$

where QPS stands for quadratic probability score, which is the counterpart of the MSFE. The range of QPS is $[0, 2]$, with 0 for perfect accuracy. A similar loss function that assigns more weight to larger forecast errors is the log probability score,

$$LPS = -\frac{1}{T}\sum_{t=1}^{T} \left((1 - R_t)\log(1 - P_t) + R_t \log P_t\right). \tag{90}$$

The range of LPS is $[0, \infty]$, with 0 for perfect accuracy.

Furthermore, Stock and Watson (1992) regressed $R_{t+k} - CRI_{t+k|t}$, i.e. the difference of their indicator of recession and the composite recession index, on available information in period t, namely

$$R_{t+k} - CRI_{t+k|t} = z_t\beta + e_t, \tag{91}$$

where the regressors in z_t are indicators included or excluded in SW's CLI. The error term in the above regression is heteroskedastic, because of the discrete nature of R_t, and serially correlated, because of the k-period ahead forecast horizon. Yet, robust t- and F-statistics can be used to test the hypothesis of interest, $\beta = 0$, that is associated with correct model specification when z_t contains indicators included in the CLI, or with an efficient use of the information in the construction of the recession forecast when z_t contains indicators excluded from the CLI. Of course, the model in (91) can be also adopted when the dependent variable is a growth rate forecast error.

If the CRI or any probability of recession are transformed into a binary indicator, S_t, by choosing a threshold such that if the probability of recession increases beyond it then the indicator is assigned a value of one, the estimation method for the regression in (91) should be changed, since the dependent variable becomes discrete. In this case, a logistic or probit regression with appropriate corrections for the standard errors of the estimated coefficients would suit.

Contingency tables can be also used for a descriptive evaluation of the methodology in the case of binary forecasts and outcomes. They provide a summary of the percentage of correct predictions, missed signals (no prediction of slowdown when it takes place), and false alarms (prediction of slowdown when it does not take place). A more formal assessment can be based on a concordance index, defined as

$$I_{RS} = \frac{1}{T}\sum_{t=1}^{T}\left[R_t S_t + (1-S_t)(1-R_t)\right], \tag{92}$$

with values in the interval $[0,1]$, and 1 for perfect concordance. Under the assumption that S_t and R_t are independent, the estimate of the expected value of the concordance index is $2\overline{SR} = 1 - \overline{R} - \overline{S}$, where $\overline{R}$ and $\overline{S}$ are the averages of R_t and S_t. Subtracting this quantity from I_{RS} yields the mean-corrected concordance index (Harding and Pagan (2001, 2002)):

$$I_{RS}^{*} = 2\frac{1}{T}\sum_{t=1}^{T}(S_t - \overline{S})(R_t - \overline{R}). \tag{93}$$

AMP showed that under the null hypothesis of independence of S_t and R_t

$$T^{1/2} I_{RS}^{*} \to N(0, 4\sigma^2), \quad \sigma^2 = \gamma_R(0)\gamma_S(0) + 2\sum_{\tau=1}^{\infty}\gamma_R(\tau)\gamma_S(\tau), \tag{94}$$

where $\gamma_S(\tau) = E[(S_t - E(S_t))(S_{t-\tau} - E(S_t))]$ and $\gamma_S(\tau)$ is defined accordingly. A consistent estimator of σ^2 is

$$\hat{\sigma}^2 = \hat{\gamma}_R(0)\hat{\gamma}_S(0) + 2\sum_{\tau=1}^{l}\left(1 - \frac{\tau}{T}\right)\hat{\gamma}_R(\tau)\hat{\gamma}_S(\tau), \tag{95}$$

where l is the truncation parameter and $\hat{\gamma}_R(\tau)$ and $\hat{\gamma}_S(\tau)$ are the sample counterparts of $\gamma_R(\tau)$ and $\gamma_S(\tau)$. As an alternative, Harding and Pagan (2001, 2002) proposed to regress R_t on S_t, and use a robust t-test to evaluate the significance of S_t.

Notice that since the predictive performance of the leading indicators can vary over expansions and recessions, and/or near turning points, it can be worth to provide a separate evaluation of the models and the indicators over these subperiods, using any of the methods mentioned so far. The comparison should be also conducted at different forecast horizons, since the ability to provide early warnings is another important property for a leading indicator, though difficult to be formally assessed in a statistical framework.

A final comment concerns the choice of the loss function, that in all the forecast evaluation criteria considered so far is symmetric. Yet, when forecasting growth or a recession indicator typically the losses are greater in case of a missed signal than for a false alarm, for example because policy makers or firms cannot take timely counteracting measures. Moreover, false alarms can be due to the implementation of timely and effective policies as a reaction to the information in the leading indicators, or can signal major slowdowns that do not turn into recessions but can be of practical policy relevance. These considerations suggest that an asymmetric loss function could be a more proper choice, and in such a case using the methods summarized so far to evaluate a leading indicator based forecast or rank competing forecasts can be misleading. For example, a model can produce a higher loss than another model even if the former has a lower MSFE or MAE, the best forecast can be biased, or an indicator can be significant in (91) without reducing the loss, see e.g. Artis and Marcellino (2001), Elliott, Komunjer and Timmermann (2003), Patton and Timmermann (2003), and Granger and Machina (2004) for an overview. More generally, the construction itself of the leading indicators and their inclusion in forecasting models should be driven by the loss function and, in case, take its asymmetry into proper account.

5.2 Examples

To illustrate the methodology for model evaluation discussed in the previous subsection, and the use of some of the models reviewed in Section 4, we now consider four empirical examples.

The first application focuses on the use of linear models for the (one-month symmetric percent changes of the) CCI_{CB} and the CLI_{CB}. We focus on the following six specifications. A bivariate VAR for the CCI_{CB} and the CLI_{CB}, as in equation (37). A univariate AR for the CCI_{CB}. A bivariate ECM for the CCI_{CB} and the CLI_{CB}, as in equation (42), where one cointegrating vector is imposed and its coefficient recursively estimated. A VAR for the four components of the CCI_{CB} and the CLI_{CB}, as in equation (32). A VAR for the CCI_{CB} and the ten components of the CLI_{CB}. Finally, a VAR for the four components of the CCI_{CB} and the ten components of the CLI_{CB}, as in equation (24). Notice that most of these models are non-nested, except for the AR which is nested in some of the VARs, and for the bivariate VAR which is nested in the ECM.

The models are compared on the basis of their forecasting performance one and six month ahead over the period 1989:1-2003:12, which includes the two recessions of July 1990 - March 1991 and March 2001 - November 2001. The forecasts are computed recursively with the first estimation sample being 1959:1-1988:12 for one step ahead forecasts and 1959:1-1988:6 for six step ahead forecasts, using the final release of the indexes and their components. While the latter choice can bias the evaluation towards the usefulness of the leading indicators, this is not a major problem when the forecasting comparison excludes the 70s and 80s and when, as in our case, the interest focuses on the comparison of alternative models for the same vintage of data, see the next subsection for details. The lag length is chosen by BIC over the full sample. Recursive BIC selects smaller models for the initial samples, but their forecasting performance is slightly worse. The forecasts are computed using both the standard iterated method, and dynamic estimation (as described in equation (28)).

The comparison is based on the MSE and MAE relative to the bivariate VAR for the CCI_{CB} and the CLI_{CB}. The Diebold and Mariano (1995) test for the statistical significance of the loss differentials is also computed. The results are reported in the upper panel of Table 6.

Five comments can be made. First, the simple AR model performs very well, there are some very minor gains from the VAR only six step ahead. This finding indicates that the lagged behaviour of the CCI_{CB} contains useful information that should be included in a leading index. Second, taking cointegration into account does not improve the forecasting performance. Third, forecasting the four components of the CCI_{CB} and then aggregating the forecasts, as in equation (34), decreases the MSE at both horizons, and the difference with respect to the bivariate VAR is significant one-step ahead. Fourth, disaggregation of the CLI_{CB} into its components is not useful, likely because of the resulting extensive parameterization of the VAR and the related increased estimation uncertainty. Finally, the ranking of iterated forecasts and dynamic estimation is not clear cut, but for the best performing VAR using the four components of the CCI_{CB} the standard iterated method decreases both the MSE and the MAE by about 10%.

In the middle and lower panels of Table 6 the comparison is repeated for, respectively, recessionary and expansionary periods. The most striking result is the major improvement of the ECM during recessions, for both forecast horizons. Yet, this finding should be interpreted with care since it is based on 18 observations only.

The second empirical example replicates and updates the analysis of Hamilton and Perez-Quiros (1996). They compared univariate and bivariate models, with and without Markov switching, for predicting one step ahead the turning points of (quarterly) GNP using the CLI_{CB} as a leading indicator, named CLI_{DOC} at that time. They found a minor role for switching (and for the use of real time data rather than final revisions), and instead a positive role for cointegration. Our first example highlighted that cointegration is not that relevant for forecasting during most of the recent period, and we wonder whether the role of switching has also changed. We use monthly data on the CCI_{CB} and the CLI_{CB}, with the same estimation and forecast sample as in the previous example. The turning point probabilities for the linear models are computed by simulations, as described at the end of Section 4.1, using a two consecutive negative growth rule to identify recessions. For the MS we use the filtered recession probabilities. We also add to the comparison a probit model where the NBER based expansion/recession indicator is regressed on six lags of the CLI_{CB}. The NBER based expansion/recession indicator is also the target for the linear and MS based forecasts, as in Hamilton and Perez-Quiros (1996).

In Table 7 we report the MSE and MAE for each model relative to the probit, where the MSE is just a linear transformation of the QPS criterion of Diebold and Rudebusch (1989), and the Diebold and Mariano (1995) test for the statistical significance of the loss differentials. The results indicate a clear preference for the bivariate MS model, with the probit a far second best, notwithstanding its direct use of the target series as dependent variable. The turning point probabilities for the five models are graphed in Figure 6, together with the NBER dated recessions (shaded areas). The figure highlights that the probit model misses completely the 2001 recession, while both MS models indicate it, and also provide sharper signals for the 1990-91 recession. Yet, the univariate MS model also gives several false alarms.

Our third empirical application is a more detailed analysis of the probit model. In particular, we consider whether the other composite leading indexes discussed in Section 3.3, the CLI_{ECRI}, CLI_{OECD}, and CLI_{SW}, or the three-month ten-year spread on the treasury bill rates have a better predictive performance than the CLI_{CB}. The estimation and forecasting sample is as in the first empirical example, and the specification of the probit models is as in the second example, namely, six lags of each CLI are used as regressors (more specifically, the symmetric one month percentage changes for CLI_{CB} and the one month growth rates for the other $CLIs$). We also consider a sixth probit model where three lags of each of the five indicators are included as regressors.

From Table 8, the model with the five indexes is clearly favoured for one-step ahead turning point forecasts of the NBER based expansion/recession indicator, with large and significant gains with respect to the benchmark, which is based on the CLI_{CB}. The second best is the ECRI indicator, followed by OECD and SW. Repeating the analysis for six month ahead forecasts, the gap across models shrinks, the term spread becomes the first or second best (depending on the use of MSE or MAE), and the combination of the five indexes remains a good choice. Moreover, the models based on these variables (and also those using the ECRI and OECD indexes) provided early warnings for both recessions in the sample, see Figures 7 and 8.

The final empirical example we discuss evaluates the role of forecast combination as a tool for enhancing the predictive performance. In particular, we combine together the forecasts we have considered in each of the three previous examples, using either equal weights or the inverse of the MSEs obtained over the training sample 1985:1-1988:12. The results are reported in Table 9.

In the case of forecasts of the growth rate of the CCI_{CB}, upper panel, the pooled forecasts outperform most models but are slightly worse than the best performing single model, the VAR with the CLI_{CB} and the four components of the CCI_{CB} (compare with Table 6). The two forecast weighting schemes produce virtually identical results. For NBER turning point prediction, middle panel of Table 9, pooling linear and MS models cannot beat the best performing bivariate MS model (compare with Table 7), even when using the better performing equal weights for pooling or adding the probit model with the CLI_{CB} index as regressor into the forecast combination. Finally, also in the case of probit forecasts for the NBER turning points, lower panel of Table 9, a single model performs better than the pooled forecast for both one and six month horizons (compare Table 8), and equal weights slightly outperforms MSE based wieghts for pooling.

5.3 Review of the recent literature on the performance of leading indicators

Four main strands of research can be identified in the recent literature on the evaluation of the performance of leading indicators. First, the consequences of the use of real time information on the composite leading index and its components rather than the final releases. Second, the assessment of the relative performance of the new models for the coincident-leading indicators. Third, the evaluation of financial variables as leading indicators. Finally, the analysis of the behavior of the leading indicators during the two most recent US recessions as dated by the NBER, namely, July 1990 - March 1991 and March 2001 - November 2001 (see e.g. McNees (1991) for results on the previous recessions). We now review in turn the main contributions in each field, grouping together the first two.

5.3.1 The performance of the new models with real time data

The importance of using real time data rather than final releases when evaluating the performance of the composite leading indicators was emphasized by Diebold and Rudebusch (1991a, 1991b). The rationale is that the composite indexes are periodically revised because of a variety of reasons including changes in data availability, timing or definition, modifications in the standardization factors, but also the past tracking performance of the index or some of its components, see Diebold and Rudebusch (1988), Swanson, Ghysels and Callan (1998) for an assessment of the revision process for the DOC-CB CLI, and Croushore (2004) for an updated overview on the use of real time data when forecasting. Therefore, an assessment of the usefulness of a composite leading index, even in a pseudo-real time framework but using the final release of the data, can yield biased results.

Diebold and Rudebusch (1991b) estimated a linear dynamic model for IP and the CLI, using dynamic estimation, and evaluated the marginal predictive content of the CLI in sample and recursively out of sample (for 1969-1988) using both finally and first released data for the CLI. While in the first two cases inclusion of the CLI in the model systematically reduces the MSFE, in the third one the results are not clear cut and depend on the lag-length and the forecast horizon. A similar finding emerges using the CCI instead of IP as the target variable, and when the Neftci's (1982) algorithm is adopted to predict turning points in IP (Diebold and Rudebusch (1991a)). Instead, using a MS model for predicting turning points, Lahiri and Wang (1994) found the results to be rather robust to the use of historical or revised data on the DOC CLI.

Filardo (1999) analyzed the performance of simple rules of thumb applied to the CLI_{CB} and of the recession probabilities computed using Neftci' (1982) formula, a linear model, a probit model, and SW's CRI, using both final and first released data over the period 1977-1998. Overall, rules of thumb and the Neftci's formula applied to the CLI_{CB} performed poorly, better with ex-post data; probit and linear models were robust to the adoption of the real-time data, because of the use of mostly financial variables as regressors, while SW's CRI was not evaluated in real time. Since the models were not directly compared on the same grounds, a ranking is not feasible but, overall, the results point towards the importance of using real-time data for the CLI also over a different and more recent sample than Diebold and Rudebusch (1991a, 1991b).

Hamilton and Perez-Quiros (1996) evaluated the usefulness of the DOC-CB CLI using linear and MS VARs, with and without cointegration, finding that the best model for predicting GDP growth and turning points over the period 1975-1993 is the linear VAR (cointegration matters in sample but not out of sample), and in this framework the CLI appears to have predictive content also with real-time data. A similar conclusion emerged from the analysis of Camacho and Perez-Quiros (2002) for the period 1972-1998, even though they found that non-linearity matters, the MS model was the best in and out of sample. Even better is a combination of the MS model with the non-parametric forecast described in Section 4.5.

A few studies compared the models described in Section 4 using the final release of the data. Notice that is less problematic in comparative analyses than in single model evaluation since all the methods can be expected to be equally advantaged. Layton and Katsuura (2001) considered logit and probit models, and a Filardo (1994) type time-varying (static) MS model, using the ECRI coincident and leading indexes. The latter model performed best in a pseudo real time evaluation exercise over the period 1979-1999, and was found to be quite useful in dating the business cycle in Layton (1998), confirming the findings in Filardo (1994). Instead, Birchenall et al. (1999) found more support for the probit model than for the MS specification.

5.3.2 Financial variables as leading indicators

Though financial variables have a long history as leading indicators, e.g. Mitchell and Burns (1938) included the Dow Jones composite index of stock prices in their list of leading indicators for the US economy, a systematic evaluation of their forecasting performance started much later, in the '80s, and since then attracted increased attention.

Stock and Watson (2003b) reviewed over 90 articles dealing with the usefulness of financial indicators for predicting output growth (and inflation), and we refer to them and to Kozicki (1997) and Dotsey (1998) for details on single studies. They also provided their own evaluation using several indicators for the G7 countries and, on the basis of the survey and of their results, concluded that some asset prices have significant predictive content at some times in some countries, but it is not possible to find a single indicator with a consistently good performance for all countries and time periods. While pooling provided a partial solution to the instability problem, Stock and Watson (2003a) suggested that "... the challenge is to develop methods better geared to the intermittent and evolving nature of these predictive relations" (p. 4).

The evidence reported in the previous and next subsections indeed points towards the usefulness of models with time-varying parameters, and also confirms the necessity of a careful choice of the financial variables to be used as leading indicators and of a continuos monitoring of their performance. A rapid survey of the literature on the interest rate spreads provides a clear and valuable illustration and clarification for this statement.

As mentioned in Section 3, Stock and Watson (1989) included two spreads into their CLI, a paper-bill spread (the difference between the 6-month commercial paper rate and the 6-month Treasury bill rate) and a term spread (the difference between the 10-year and the 1-year Treasury bond rates.

The paper-bill spread tends to widen before a recession reflecting expectations of business bankruptcies, corporations' growing cash requirements near the peak of the business cycle, and tighter monetary policy (the paper rate rises because banks deny loans due to the restricted growth of bank reserves, so that potential borrowers seek funds in the commercial paper marker). Yet, the paper bill-spread could also change for other reasons unrelated to the business cycle, such as changes in the Treasury's debt management policy, or foreign central banks interventions in the exchange market since a large amount of their reserves in dollars are invested in Treasury bills, see e.g. Friedman and Kutnner (1998), who found these reasons capable of explaining the bad leading performance of the paper-bill spread for the 1990-91 recession, combined with the lack of a tighter monetary policy. The performance for the 2001 recession was also unsatisfactory, the spread was small and declining from August 2000 to the end of 2001, see also the next subsection.

The term spread has two components, expected changes in interest rates and the term premium for higher risk and/or lower liquidity. Therefore the commonly observed negative slope of the term structure prior to recession, i.e. long term rates becoming lower than short term ones, can be due either to lower expected short term rates (signaling expansionary monetary policy) or to lower term premia. Hamilton and Kim (2002) found both components to be relevant for forecasting output growth, with the former dominating at longer forecast horizons. The bad leading performance of the term spread for the 1990-91 recession is also typically attributed to the lack of a tighter monetary policy in this specific occasion. The term spread became instead negative from June 2000 through March 2001, anticipating the recession of 2001, but the magnitude was so small by historical standards that, for example, SW's composite leading index did not signal the recession, see also the next subsection.

Gertler and Lown (2000) suggested to use the high-yield (junk) / AAA bond spread as a leading indicator, since it is less sensitive to monetary policy and provides a good proxy for the premium for external funds, i.e., for the difference between the costs of external funds and the opportunity costs of using internal funds. The premium for external funds moves countercyclically, since during expansions the borrowers' financial position typically improves, and this further fosters the aggregate activity, see e.g. Bernanke and Gertler (1989) for a formalization of this final accelerator mechanism. Therefore, a widening high-yield spread signals a deterioration of economic conditions. Gertler and Lown (2000) found that after the mid 80's the high-yield spread had a better forecasting performance than both the paper-bill and the term spreads for the US GDP growth, providing also a warning for the 1990-91 recession. Yet, as for the paper-bill spread, the high-yield spread can also change for reasons unrelated with the business cycle, such as confidence crises in emerging markets. In particular, Duca (1999) indicated that the widening of the spread prior to the 1990-91 recession could be an accidental event related with the thrift crisis and the associated sale of junk bonds in an illiquid market.

A related question of interest is whether it is better to use a financial indicator in isolation or as a component of a composite index. Estrella and Mishkin (1998) ran probit regressions using the term-spread, the CLI_{CB}, the CLI_{SW}, and some of their components, concluding that both in sample and out of sample the spread yields the largest forecasting gains. Moreover, addition of other regressors is in general harmful, except for the NYSE index returns. Similar conclusions emerged from the analysis in Dueker (1997), who also used more complicated versions of the probit model, allowing for dynamics and Markov switching parameters. Qi (2001) also obtained a similar finding using the neural network model described in Section 4.5. The CLI_{SW} was best at 1-quarter forecast horizon, but the term spread at 2- to 6-quarter horizon. Yet, she also detected substantial instability of the results over different decades, namely, the '70s, '80s, and '90s. Estrella, Rodrigues and Schich (2000) also found some instability for the US, more so when the dependent variable is the GDP growth rate than when it is a binary expansion/recession indicator.

Chauvet and Potter (2001a) detected substantial instability also in the probit model when it is estimated with the Gibbs sampler. Moreover, the date of the break has a major role in determining the predictive performance of the spread, for example the probability of a future recession are about 45% in December 2000 when no break is assumed but increase to 90% imposing a break in 1984. Unfortunately, there is considerable uncertainty about the break date, so that the posterior mean probability of recession across all break dates is 32% with a 95% interval covering basically the whole $[0, 1]$ interval. Chauvet and Potter (2001b) extended the basic probit model to allow for parameter instability, using a time-varying specification, and also for autocorrelated errors. Though the more complicated models performed better, along the lines of Dueker (1997), they provided a weaker signal of recession in 2001 in a real-time evaluation exercise.

Finally, positive results on the leading properties of the term spread and other financial variables for other countries were reported, e.g. by Davis and Henry (1994), Davis and Fagan (1997), Estrella and Mishkin (1997), Estrella et al. (2000), and Moneta (2003). Yet, Moneta (2003) found also for the Euro area a deterioration in the relative leading characteristics of the spread after the '80s, and an overall unsatisfactory performance in predicting the Euro area recession of the early '90s.

5.3.3 The 1990-91 and 2001 US recessions

Stock and Watson (1993) conducted a detailed analysis of possible reasons for the failure of their CRI to produce early warnings of the 1990-91 recession. They could not detect any signs of model failure or mis-specification and therefore concluded that the major problem was the peculiar origin of this recession compared with its predecessors, namely, a deterioration in the expectations climate followed by a drop in consumption. In such a case, the treasury bill yield curve, exchange rates, and partly IP provided wrong signals. Only three other leading indicators in their set gave moderate negative signals, part-time work, building permits and unfilled orders, but they were not sufficiently strong to offset the other indicators.

Phillips (2003) compared the performance of the CRI_{SW}, and of the CLI_{CB} and the term spread, transformed into probabilities of recession using Neftci's (1982) formula, for forecasting the 1990-91 recession using real time data. He found that that the CLI_{CB} produced the best results. Moreover, the SW's index modified to allow for longer lags on the term and quality spreads worked better in sample but not for this recession.

Chauvet (1998) also used a real time dataset to produce recession forecasts from her dynamic MS factor model, and found that the filtered probability of recession peaked beyond 0.5 already at the beginning of 1990 and then in May of that year.

Filardo and Gordon (1999) contrasted a linear VAR model, a MS model with time-varying parameters, the SW's model, and a MS factor model with time-varying parameters, along the lines of Chauvet (1998). All models were estimated using Gibbs sampling techniques, and compared on the basis of the marginalized likelihoods and Bayes factors in 1990, as suggested by Geweke (1994), since these quantities are easily computed as a by-product of the estimation. They found that all models performed comparatively over the period January-June, but in the second part of the year, when the recession started, the MS model was ranked first, the VAR second, and the factor model third, with only minor differences between the two versions.

Filardo (2002), using the same models as in Filardo (1999) found that the two-month rule on the CLI_{CB} worked well in predicting the 2001 recession, but sent several false alarms in the '90s. A probit model with a 3-month forecast horizon and the term spread, corporate spread, S&P500 returns and the CLI_{CB} as regressors also worked well, predicting the beginning of the recession in January 2001 using a 50% rule. Instead, the CRI_{SW} did not perform well using a 50% rule, while SW's $CRI - C$ (contemporaneous) worked better but was subject to large revisions.

Stock and Watson (2003a) analyzed in details the reasons for the poor performance of the CRI, concluding that is was mostly due to the particular origin of the recession (coming from the decline in stock prices and business investment), which is not properly reflected by most of the indicators in their CRI. In particular, the best indicators for the GDP growth rate were the term spread, the short term interest rate, the junk bond spread, stock prices, and new claims for unemployment. Notice that most of these variables are included in Filardo's (2002) probit models. Moreover, they found that pooled forecasts worked well, but less well than some single indicators in the list reported above.

Dueker (2003) found that his Qual-VAR predicted the timing of the 2001 recession quite well relative to the professional forecasters, while the evidence in Dueker and Welshe (2001) is more mixed. Dueker (2002) noticed that a MS-probit model with the CLI_{CB} as regressor worked also rather well in this occasion, providing a 6-month warning of the beginning of the recession (but not in the case of the previous recession).

Overall, some differences in the ranking of models and usefulness of the leading indicators emerged because of the choice of the specific coincident and leading variables, sample period, criteria of evaluation, etc. Yet, a few findings are rather robust. First, indicator selection and combination methods are important, and there is hardly a one fits all choice, even though financial variables and the equal weighted CLI_{CB} seem to have a good average performance. Second, the model that relates coincident and leading indicators also matters, and a MS feature is systematically helpful. Finally, pooling the forecasts produced good results whenever applied, even though there is only a limited evidence as far as turning points are concerned.

6 What have we learned?

The experience of the last two recessions in the US confirmed that these are difficult events to predict, because the generating shocks and their propagation mechanism change from time to time, and there is a very limited sample to fit the more and more complex models that try to capture these time-varying features. Nonetheless, the recent literature on leading indicators provided several new useful insights for the prediction of growth rates and turning points of a target variable.

The first set of improvements is just in the definition of the target variable. In Section 2 we have seen that several formal procedures were developed to combine coincident indicators into a composite index, which is in general preferable to monitoring a single indicator because of its narrower coverage of the economy. In practice, the new model based $CCIs$ are very similar to the old-style equal averages of the (standardized) coincident indicators, such as the CCI_{CB}, but they provide a sounder statistical framework for the use and evaluation of the $CCIs$. More sophisticated filtering procedures were also developed to emphasize the business cycle information in a CCI, even though substantial care should be exerted in their implementation to avoid phase shifts and other distortions. New methods were also developed for dating the peaks and troughs in either the classical or the deviation cycle. They closely reproduce the NBER dating for the US and the CEPR dating for the euro area, but are more timely and can also provide a probabilistic measure of uncertainty around the dated turning points.

The second set of advances concerns the construction of leading indicators. While there was general agreement on the characteristics of a good leading indicator, such as consistent timing or conformity to the general business cycle, in Section 3 we have seen that there are now better methods to formally test the presence of these characteristics and assess their extent. Moreover, there were several developments in the construction of the composite leading indexes, ranging from taking into explicit account data problems such as missing values or measurement error, to an even more careful variable selection relying on new economic and statistical theories, combined with sounder statistical procedures for merging the individual leading indicators into a CLI.

The third, and perhaps most important, set of enhancements is in the use of the leading indicators. In Section 4 we have seen that simple rules to transform a CLI into a turning point forecasts have been substituted with sophisticated non-linear and time-varying models for the joint evolution of the coincident and leading indicators. Moreover, mainly using simulation-based techniques, it is now rather easy to use a model to produce both point and probability and duration forecasts.

The final set of improvements is in the evaluation of leading indicators. In Section 5 we have seen that formal statistical methods are now available to assess the forecasting performance of leading indicators, possibly combined with the use of real time data to prevent biased favorable results due to revisions in the composition of the $CLIs$. Moreover, an overview of the forecasting performance over the two most recent recessions in the US has provided some evidence in favor of the forecasting capabilities of $CLIs$, in particular when simple weighting procedures are applied to a rather large set of indicators, combined with sophisticated models for the resulting CLI and the target variable.

Notwithstanding the substantial progress in the recent years, there is still considerable scope for research in this area. For example, it might be useful to achieve a stronger consensus on the choice of the target variable, and in particular on whether the classical cycle is really the target of interest or a deviation cycle could provide more useful information. The collection of higher quality monthly series and the development of better methods to handle data irregularities also deserve attention. But the crucial element remains the selection of the leading variables, and of the weighting scheme for their combination into a CLI. Both choices should be made endogenous and frequently updated to react to the changing shocks that hit the economy, and further progress is required in this area. Forecast pooling could provide an easier method to obtain more robust predictions, but very limited evidence is available for turning point and duration forecasts. It is also worth mentioning that while in this chapter we have focused on real activity as the target variable, other choices are possible such as inflation or a stock market index, see e.g. the contributions in Lahiri and Moore (1991), and most of the developments we have surveyed could be usefully applied in these related contexts.

To conclude, most of what we have learned in the recent period about leading indicators builds upon ideas originally developed in a set of papers all published in 1989. 1989 is also the year of the fall of the Berlin's wall, which opened the way to the transition of most previously communist countries towards the market economy. Since

> "Business cycles are a type of fluctuation found in the aggregate economic activity of nations that organize their work mainly in business enterprises", Burns and Mitchell (1946, p.3),

leading indicators have an even wider market now, and the construction of indicators for these new members of the market economy is the final item we would include in the "to do list".

References

Albert, J., and S. Chib (1993a), "Bayesian analysis via Gibbs sampling of autoregressive time series subject to Markov mean and variance shifts", *Journal of Business and Economic Statistics* 11: 1-15.

Albert, J. and S. Chib (1993b), "Bayesian analysis of binary and polychotomous response data", *Journal of the American Statistical Association* 88: 669-679.

Altissimo, F., Bassanetti, A., Cristadoro, R., Forni, M., Lippi, M., Reichlin L. and Veronese, G., (2001), " EuroCoin: A real time coincident indicator of the Euro area business cycle", CEPR Working Paper 3108

Anderson, H. M. and F. Vahid (2001), "Predicting the probability of a recession with nonlinear autoregressive leading-indicator models", *Macroeconomic Dynamics* 5: 482-505.

Artis, M.J., D. Osborn, .R., Bladen-Hovell, G. Smith and W. Zhang (1995), "Turning Point Prediction for the UK Using CSO Leading Indicators", *Oxford Economic Papers*, 4, July

Artis, M. J., Banerjee, A. and M. Marcellino(2004), "Factor forecasts for the UK", *Journal of Forecasting*, forthcoming.

Artis, M. J., Canova, F., Gali, J., Giavazzi, F., Portes, R., Reichlin, L., Uhlig, H. and P. Weil (2003), "Business cycle dating committee of the Centre for Economic Policy Research", CEPR.

Artis, M. J., Galvao, A. B. and M. Marcellino (2003), "The transmission mechanism in a changing world", CEPR Working Paper no. 4014.

Artis, M. J., Krolzig, H.-M, and J. Toro (2003), "The European business cycle", CEPR Discussion Paper no. 2242.

Artis, M.J., and M. Marcellino (2001), "Fiscal forecasting: the track record of IMF, OECD and EC", *Econometrics Journal* 4: s20-s36.

Artis, M.J., M. Marcellino and T. Proietti (2003), "Dating the Euro area business cycle", *Oxford Bulletin of Economics and Statistics*, forthcoming.

Auerbach, A. J. (1982), "The index of leading indicators: "Measurement without theory' thirty-five years later", *Review of Economics and Statistics* 64(4): 589-595.

Bai, J. (2003), "Inferential Theory for Factor Models of Large Dimensions," *Econometrica*, 71, 135-171

Bai, J. and S. Ng (2002), "Determining the number of factors in approximate factor models", *Econometrica* 70(1): 191-221.

Bai, J., and S. Ng (2003), "Confidence intervals for diffusion index forecasts with a large number of preditors", Working Paper, University of Michigan.

Banerjee, A. N. and M. Marcellino (2003), "Are there any reliable leading indicators for US inflation and GDP growth?", Economics Working Papers ECO2002/21, European University Institute.

Banerjee, A. N., Marcellino, M., and I. Masten (2003), "Leading indicators for Euro area inflation and GDP growth", CEPR Discussion Paper no. 3893.

Bates, J. M. and C. W. J. Granger (1969), "The combination of forecasts", *Operations Research Quarterly* 20: 451-468.

Baxter, M. and R.G. King (1999), "Measuring Business Cycles: Approximate Band-Pass Filters for Macroeconomic Time Series," *Review of Economics and Statistics*, 81(4): 575-93.

Bernanke, B. S. and M. L. Gertler (1989), "Agency costs, net worth, and business fluctuations", *American Economic Review* 79: 14-31.

Bernard, H., and S. Gerlach (1998), "Does the term structure predict recessions? The international evidence", *International Journal of Finance and Economics* 3: 195-215.

Birchenhall, C. R., Jessen, H., Osborn, D. R., and P. Simpson (1999), "Predicting US business cycle regimes", *Journal of Business and Economic Statistics* 17(3): 313-323.

Birchenhall, C. R., Osborn, D. R. and M. Sensier (2001), "Predicting UK Business Cycle Regimes", *Scottish Journal of Political Economy*, 48(2), 179-95

Boehm, E. A. (2001), "The Contribution of Economic Indicator Analysis to Understanding and Forecasting Business Cycles," *Indian Economic Review*. Vol. 36. Pp. 1-36.

Boivin, J. and S. Ng (2003), "Are more data always better for factor analysis", NBER Working Paper no. 9829 forthcoming, *Journal of Econometrics*.

Boldin, M. D.(1994), "Dating Turning Points in the Business Cycle," *Journal of Business*, Vol. 67 (1) pp. 97-131. University of Chicago Press

Brillinger, David R. (1981), "Time series data analysis and theory", Holt, Rinehart, and Winston (New York).

Bry, G. and C. Boschan (1971), "Cyclical Analysis of Time Series: Selected Procedures and Computer Programs," Technical Paper 20, NBER, Columbia University Press

Burns, A. F. and W. C. Mitchell (1946), "Measuring business cycles", NBER Studies in Business Cycles no. 2 (New York).

Camacho, M. 2004, "Vector smooth transition regression models for US GDP and the composite index of leading indicators", *Journal of Forecasting*, forthcoming.

Camacho, M. and G. Perez-Quiros (2002), "This is what the leading indicators lead", *Journal of Applied Econometrics* 17(1): 61-80.

Camba-Mendez G., G.Kapetanios, R. J. Smith and M. R. Weale (2001), "An Automatic Leading Indicator of Economic Activity: Forecasting GDP Growth for European Countries", *Econometrics Journal* 4(1), S56-90

Canova, F. (1999), "Reference Cycle and Turning Points: A Sensitivity to Detrending and Classification Rules" *Economic Journal*, 1999, 112, 117-142.

Canova, F. (2004), Method for Applied Research, unpublished

Canova, F. and M. Ciccarelli (2001), "Forecasting and Turning Point Predictions in a Bayesian Panel VAR Model", CEPR Discussion Paper no. 2961.

Canova, F. and M. Ciccarelli (2003), "Panel Index VAR Models: Specification, Estimation, Testing and Leading Indicators ", CEPR Discussion Paper no. 4033.

Carter, C. K. and R. Kohn (1994), "On Gibbs sampling for state space models", *Biometrika* 81: 541-55.

Chauvet, M. (1998), "An econometric charcterization of business cycle dynamics with factor structure and regime switching", *International Economic Review* 39(4): 969-996.

Chauvet, M. and J.M. Piger (2003), "Identifying Business Cycle Turning Points in Real Time", *Federal Reserve Bank of St. Louis Review*, 85(2), 13-26

Chauvet, M. and S. Potter (2001a), "Predicting a recession: evidence from the yield curve in the presence of structural breaks", Working Paper (Federal Reserve Bank of New York).

Chauvet, M. and S. Potter (2001b), "Forecasting recessions: using the yield curve", Working Paper (Federal Reserve Bank of New York

Chen, C. W. S. and Lee, J. C. (1995), "Bayesian inference of threshold autoregressive models", *Journal of Time Series Analysis*, 16, 483-492.

Chin, D., Geweke, J. and P. Miller (2000), "Predicting turning points", Federal Reserve Bank of Minneapolis Staff Report no. 267.

Christiano, L. and T. Fitzgerald (2003), "The Band Pass Filter," *International Economic Review*, vol. 44, iss. 2, pp. 435-465(31)

Clark, T. E., and M. W. McCracken (2001), "Test of equal forecast accuracy and encompassing for nested models", *Journal of Econometrics* 105: 85-100.

Clements, M.P. and D. F. Hendry (1996), "Multi-step estimation for forecasting", *Oxford Bulletin of Economics and Statistics*, 58, 657-684.

Clements, M.P. and D. F. Hendry (1999), *Forecasting Non-stationary Economic Time Series*, Cambridge, Mass.: MIT Press

Corradi, V., and N. R. Swanson (2004), "Density and Interval Forecasting", in G. Elliott, C.W.J. Granger and A. Timmermann (eds.) Handbook of Economic Forecasting

Croushore, D. (2004), "Real-Time Data", in G. Elliott, C.W.J. Granger and A. Timmermann (eds.) Handbook of Economic Forecasting

Davis, E. P., and G. Fagan (1997), "Are financial spreads useful indicators of future inflation and output growth in EU countries?", *Journal of Applied Econometrics* 12: 701-714.

Davis, E. P. and S. G. B. Henry (1994), "The use of financial spreads as indicator variables: Evidence for the United Kingdom and Germany", IMF Staff Papers 41: 517-525.

Del Negro, M. (2001), "Turn, Turn, Turn: Predicting Turning Points in Economic Activity", *Federal Reserve Bank of Atlanta Economic Review* 87

Diebold, F. X., Lee, J.-H., and G. C. Weinbach (1994), "Regime switching with time-varying transition probabilities", in C. Hargreaves, ed., Non-*stationary time-series analyses and cointegration*, Oxford University Press (Oxford): 283-302.

Diebold, F. X., and R. Mariano (1995), "Comparing predictive accuracy", *Journal of Business and Economic Statistics* 13: 253-263.

Diebold, F. X., and G. D. Rudebusch (1988), "Stochastic properties of revisions in the index of leading indicators", Proceedings of the Business and Economic Statistics Section (American Statistical Association): 712-717.

Diebold, F. X., and G. D. Rudebusch (1989), "Scoring the leading indicators", *The Journal of Business* 62(3): 369-391.

Diebold, F. X. and G. D. Rudebusch (1990), "A nonparametric investigation of duration dependence in the American business cycle", *Journal of Political Economy* 98: 596-616.

Diebold, F. X. and G. D. Rudebusch (1991a), "Turning point prediction with the composite leading index: an ex ante analysis", in Lahiri, K., and G. H. Moore, eds., *Leading economic indicators: new approaches and forecasting records*, Cambridge University Press (Cambridge, UK): 231-256.

Diebold, F. X. and G. D. Rudebusch (1991b), "Forecasting output with composite leading index: a real-time analysis", *Journal of the American Statistical Association* 86(415): 603-610.

Diebold, F. X. and G. D. Rudebusch (1996), "Measuring business cycles: a modern perspective", *The Review of Economics and Statistics* 78(1): 67-77.

Diebold, F. X., Rudebusch, G. D. and D. E. Sichel (1993), "Further evidence on business cycle duration dependence", in: Stock, J. H., and M. W. Watson, eds., *Business Cycles, Indicators, and Forecasting*, The University of Chicago Press (Chicago): 255-280.

van Dijk, D., Terasvirta, T. and P. H. Franses (2002), "Smooth transition autoregressive models - A survey of recent developments", Econometric Reviews, 21, 1:47.

Doan, T., R. Litterman, and C.Sims (1984), "Forecasting and Conditional Projection Using Realistic Prior Distributions", *Econometric Review*, 3, 1-100.

Dotsey, M. (1998), "The predictive content of the interest rate term spread for future economic growth", *Federal Reserve Bank of Richmond Economic Quarterly* 84(3): 31-51.

Duca, J. V. (1999), "What credit market indicators tell us", *Federal Reserve Bank of Dallas Economic and Financial Review* (1999:Q3): 2-13.

Dueker, M. J. (1997), "Strengthening the case for the yield curve as a predictor of US recessions", *Federal Reserve Bank of St. Louis Review* no. 79(2): 41-51.

Dueker, M. J. (2002), "Regime-dependent recession forecasts and the 2001 recession", *Review (Federal Reserve Bank of St. Louis).*

Dueker, M. J. (2003), "Dynamic forecasts of qualitative variables: a Qual VAR model of US recessions", *Journal of Business and Economic Statistics*, forthcoming.

Dueker, M. J., and K. Wesche (2001), "Forecasting output with information from business cycle turning points: a qualitative variable VAR", Working Paper (Federal Reserve Bank of St. Louis).

Durland, J. M. and T. H. McCurdy (1994), "Duration-dependent transitions in a Markov model of US GNP growth", *Journal of Business and Economic Statistics* 12: 279-288.

Eichengreen, B., Watson, M. W., and R. S. Grossman (1985), "Bank rate policy under the interwar gold standard", *Economic Journal* 95: 725-745.

Elliott, G., Komunjer, I., and A. Timmermann (2003), "Estimating loss function parameters", CEPR Discussion Papers no. 3821.

Emerson, R. A. and D. F. Hendry (1996), "An evaluation of forecasting using leading indicators", *Journal of Forecasting*, 15: 271-291.

Estrella, A., and F. S. Mishkin (1997), "The predictive power of the term structure of interest rates in Europe and the United States: Implications for the European Central Bank", *European Economic Review* 41: 1375-1401.

Estrella, A. and F. S. Mishkin (1998), "Predicting US recessions: financial variables as leading indicators", *The Review of Economics and Statistics* 80(1): 45-61.

Estrella, A., Rodrigues, A. P. and S. Schich (2000), "How stable is the predictive power of the yield curve? Evidence from Germany and the United States", Federal Reserve Bank of New York Staff Report no. 113.

Fair, R. (1993), "Estimating event probabilities from macroeconomic models using stochastic simulation", in: Stock, J. H., and M. W. Watson, eds., *Business Cycles, Indicators, and Forecasting*, The University of Chicago Press (Chicago): 157-178.

Filardo, A. J. (1994), "Business cycle phases and their transitional dynamics", *Journal of Business and Economic Statistics* 12(3): 299-308.

Filardo, A. J. (1999), "How reliable are recession prediction models?", *Federal Reserve Bank of Kansas City Economic Review* 84(2): 35-55.

Filardo, A. J. (2002), "The 2001 US recession: what did recession prediction models tell us?", Manuscript (Bank for International Settlements).

Filardo, A. J. and S. F. Gordon (1994), "International Co-movements of Business Cycles", Research Working Paper 94-11, Federal Reserve Bank of Kansas

Filardo, A. J. and S. F. Gordon (1998), "Business cycle durations", *Journal of Econometrics* 85: 99-123

Filardo, A. J. and S. F. Gordon (1999), "Business cycle turning points: two empirical business cycle model approaches", in: P. Rothman, ed., *Nonlinear time series analysis of economic and financial data*, vol. 1 (Kluwer Academic Publishers), ch. 1: 1-32.

Forni, M., Hallin, M., Lippi, M. and L. Reichlin (2000), "The generalized factor model: identification and estimation", *The Review of Economics and Statistics* 82(4): 540-554.

Forni, M., Hallin, M., Lippi, M. and L. Reichlin (2001), "Coincident and leading indicators for the Euro area", *The Economic Journal* 111(May): C62-C85.

Forni, M., Hallin, M., Lippi, M. and L. Reichlin (2003a), "The generalized dynamic factor model: one-sided estimation and forecasting", revised version, mimeo.

Forni, M., Hallin, M., Lippi, M. and L. Reichlin (2003b), "Do financial variables help forecasting inflation and real activity in the Euro area?", *Journal of Monetary Economics* 50, 1243-55.

Friedman, B. M. and K. N. Kutnner (1998), "Indicator properties of the paper-bill spread: Lessons from recent experience", *The Review of Economics and Statistics* 80(1): 34-44.

Fuhrer, J. C. and G. D. Rudebusch (2002), “Estimating the Euler equation for output”, Working Papers 02-3, Federal Reserve Bank of Boston.

Fuhrer, J. C. and S.Schuh (1998), “Beyond Shocks: What Causes Business Cycles? An Overview" Beyond Shocks: What Causes Business Cycles? Proceedings from the Federal Reserve Bank of Boston Conference Series no. 42, pp. 1-31.

Gertler, M. and C. S. Lown (2000), "The information in the high yield bond spread for the business cycle: evidence and some implications", NBER Working Paper Series, no. 7549.

Geweke, J. (1977), "The dynamic factor analysis of economic time series", in: Aigner, D. J., and A. S. Goldberger, eds., *Latent variables in socio-economic models*, North Holland Publishing (Amsterdam), ch. 19.

Geweke, J. (1994), "Variable selection and model comparison in regression", in: Berger, J. O., Bernardo, J. M., Dawid, A. P., and A. F. M. Smith, eds., Proceedings of the Fifth Valencia International Meetings on Bayesian Statistics.

Geweke, J. and N. Terui (1993), “Bayesian Threshold Autoregressive Models for Nonlinear Time Series,” *Journal of Time Series Analysis*, 1993, 14, 441-455.

Geweke, J. C.H. Whiteman (2004), “Bayesian Forecasting” in preparation for *The Handbook of Economic Forecasting*, eds. G. Elliott, C.W.J. Granger, and A. Timmermann, 2006. Elsevier.

Granger, C. W. J. and M. J. Machina (2004), “Decision Theory and Forecasting” in preparation for *The Handbook of Economic Forecasting*, eds. G. Elliott, C.W.J. Granger, and A. Timmermann, 2006. Elsevier.

Granger, C. W. J. and R. Ramanathan (1984), "Improved methods of combining forecasts", *Journal of Forecasting* 3: 197-204.

Granger, C. W. J., Terasvirta, T., and H. M. Anderson (1993), "Modelling nonlinearity over the business cycle", in: Stock, J. H., and M. W. Watson, eds., *Business Cycles, Indicators, and Forecasting*, The University of Chicago Press (Chicago): 311-325.

Hall, R.E., M. Feldstein, J. Frankel, R. Gordon, C. Romer, D. Romer and V. Zarnowitz (2003),"The NBER’s Recession Dating Procedure", Business Cycle Dating Committee, National Bureau of Economic Research

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 University Press (Princeton).

Hamilton, J. D. and D. H. Kim (2002), "A re-examination of the predictability of economic activity using the yield spread", *Journal of Money, Credit and Banking* 34: 340-360.

Hamilton, J. D. and G. Perez-Quiros (1996), "What do the leading indicators lead?", *The Journal of Business* 69(1): 27-49.

Hamilton, J. D. and B. Raj (2002), "New directions in business cycle research and financial analysis", *Empirical Economics* 27(2): 149-162.

Harding, D. and A. Pagan (2001), "Dissecting the cycle", forthcoming, *Journal of Monetary Economics*.

Harding, D. and A. Pagan (2002), “Dissecting the cycle: A methodological investigation”, *Journal of Monetary Economics* 49(2): 365-381.

Haerdle, W. and P. View (1992), "Kernel regression smoothing of time series", *Journal of Time Series Analysis* 13: 209-232.

Hendry, D. F. (1995), *Dynamic Econometrics*. Oxford: Oxford University Press, 1995.

Hendry, D. F. and H.-M. Krolzig (1999), "Improving on `data mining reconsidered' by K. D. Hoover and S. J. Perez", *Econometrics Journal*, 2, 167-191

Hendry, D. F. and H.-M. Krolzig (2001), "Computer Automation of General-to-Specific Model Selection Procedures", *Journal of Economic Dynamics and Control*, 25 (6-7), 831-866.

Hodrick, R.J. and E.C. Prescott (1997), "Postwar US Business Cycles: An Empirical Investigation", *Journal of Money, Credit and Banking*, 29, 1-16.

Hornick, K., Stinchcombe, M. and H. White (1989), "Multilayed feedforward networks are universal approximators", *Neural Networks* 2: 359-366.

Johansen, S.(1988), "Statistical analysis of cointegration vectors", *Journal of Economic Dynamics and Control* 12: 231-254.

Kapetanios, G. and M. Marcellino (2003), "A Comparison of Estimation Methods for Dynamic Factor Models of Large Dimensions", Working Papers no. 489, Queen Mary, University of London, Department of Economics.

Keilis-Borok, V., Stock, J. H., Soloviev, A. and P. Mikhalev (2000), "Pre-recession pattern of six economic indicators in the USA", *Journal of Forecasting* 19: 65-80.

Kim, C.-J. (1994), "Dynamic linear models with Markov switching", Journal of Econometrics 60: 1-22.

Kim, C.-J. and C. Murray (2002), "Permanent and transitory components of recessions", *Empirical Economics* 27: 163-183.

Kim, C.-J. and C. R. Nelson (1998), "Business cycle turning points, a new coincident index, and tests of duration dependence based on a dynamic factor model with regime switching", *The Review of Economics and Statistics* 80: 188-201.

Kim, M.-J. and J.-S. Yoo (1995), "New index of coincident indicators: a multivariate Markov switching factor model approach", *Journal of Monetary Economics* 36: 607-630.

Kling, J. L. (1987), "Predicting the turning points of business and economic time series", *Journal of Business* 60(2): 201-238.

Koch, P. D. and R. H. Rasche (1988), "An examination of the commerce department leading-indicator approach", *Journal of Business and Economic Statistics* 6(2): 167-187.

Koopmans, T. C. (1947), "Measurement without theory", *Review of Economics and Statistics* 29: 161-179.

Kozicki, S. (1997), "Predicting real growth and inflation with the yield spread", *Federal Reserve Bank of Kansas City Economic Review*, Fourth Quarter.

Krolzig, H.-M. (1997), "Markov switching vector autoregressions. Modelling, statistical inference and application to business cycle analysis", Springer (Berlin).

Krolzig, H.-M. (2004), "Predicting Markov-switching vector autoregressive processes", *Journal of Forecasting*, forthcoming.

Krolzig, H.-M., M. Marcellino and Mizon (2002), "A Markov-switching vector equilibrium correction model of the UK labour market", *Empirical Economics* 27(2): 233-254.

Lahiri, K. and G.H. Moore (1991), *Leading economic indicators: New approaches and forecasting records*, Cambridge University Press, Cambridge

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

Layton, A. P. (1996), "Dating and Predicting Phase Changes in the US Business Cycle", *International Journal of Forecasting*, 12, 417-28

Layton, A. P. (1998), "A further test of the leading indicators on the probability of US business cycle phase shifts", *International Journal of Forecasting* 14: 63-70

Layton, A. P. and M. Katsuura (2001), "Comparison of regime switching, probit and logit models in dating and forecasting US business cycles", *International Journal of Forecasting* 17: 403-417.

Li, D.T. and J.H. Dorfman (1996), "Predicting turning points through the integration of multiple models", *Journal of Business and Economic Statistics* 14(4): 421-428.

Lieli, R. P. (2004), 2003?? "A flexible framework for predicting binary variables", Job Market Paper (UC San Diego).

Lindgren, G. (1978), "Markov regime models for mixed distributions and switching regressions", *Scandinavian Journal of Statistics* 5: 81-91.

Lubrano, M., (1995), "Bayesian Tests for Co-Integration in the Case of Structural Breaks: An Application to the Analysis of Wages Moderation in France", G.R.E.Q.A.M. 95a19, Universite Aix-Marseille III

Lütkepohl, H. (1987), *Forecasting Aggregated Vector ARMA Processes*, Springer-Verlag, Berlin.

Lütkepohl, H. (2004), "Forecasting with VARMA Models", in G. Elliott, C.W.J. Granger and A. Timmermann (eds.) Handbook of Economic Forecasting

Marcellino M. (2004a), "Forecast pooling for short time series of macroeconomic variables", *Oxford Bulletin of Economics and Statistics*, 66, 91-112.

Marcellino, M. (2004b), "Forecasting EMU macroeconomic variables", *International Journal of Forecasting*, forthcoming.

Marcellino, M. and G. Mizon (2003), "Progressive Modelling: Encompassing and hypothesis testing", Oxford: Oxford University Press, forthcoming.

Marcellino, M., Stock, J. H. and M. W. Watson (2003), "Macroeconomic forecasting in the Euro area: country specific versus Euro wide information", *European Economic Review*, 47, 1-18.

Marcellino, M., Stock, J. H. and M. W. Watson (2004), "A Comparison of Direct and Iterated AR Methods for Forecasting Macroeconomic Series h-Steps Ahead", mimeo.

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

McGuckin, R.H., A. Ozyildirimand V. Zarnowitz (2003), "A More Timely and Useful index of Leading Indicators", Working Paper

McNees, S.K. (1991), "Forecasting cyclical turning points: the record in the past three recessions", in *Leading Economic Indicators: New approaches and forecasting records*, (eds.) K. Lahiri, G. H. Moore:Cambridge University Press

Mitchell, W. and A. F. Burns (1938), "Statistical indicators of cyclical revivals", NBER (New York), reprinted in: G. H. Moore, ed., (1961), *Business cycle indicators*, Princeton University Press (Princeton), ch. 6.

Moneta, F. (2003), "Does the yield spread predict recessions in the Euro area?", ECB Working Paper no. 294.

Moore, G. H. and J. Shiskin (1967), "Indicators of business expansions and contractions", NBER Occasional Paper no. 103.

Moore, G. H. (1983), *Business Cycles, Inflation and Forecasting*, 2nd ed.,

Moore, G. H. and V. Zarnowitz (1986), "The Development and Role of the NBER's Business Cycle Chronologies," in Robert J. Gordon, ed., *The American Business Cycle: Continuity and Change*, Chicago: University of Chicago Press, pp. 735-779

Neftci, S. N. (1982), "Optimal prediction of cyclical downturns", *Journal of Economic Dynamics and Control* 4: 225-241.

Niemira, M.P. and Klein, P.A (1994), Forecasting Financial and Economic Cycles, John Wiley and Sons Ed.

Osborn, D., M. Sensier and P.W.Simpson (2001), "Forecasting UK Industrial Production over the Business Cycle", *Journal of Forecasting*, 20(6), 405-24

Otrok, C. and C.H. Whiteman (1998), "Bayesian Leading Indicators: Measuring and Predicting Economic Conditions in Iowa", *International Economic Review* 39 , 997-1014.

Patton, A. and A. Timmermann (2003), “Properties of Optimal Forecasts”, mimeo.

Phillips, K. R. 2003, "The composite index of leading economic indicators: a comparison of approaches", Working Paper, Federal Reserve Bank of Dallas

Potter, S. M. (1995), "A nonlinear approach to US GNP", *Journal of Applied Econometrics* 10: 109-125.

Proietti, T. and F. Moauro (2003), "Dynamic Factor Analysis with Nonlinear Temporal Aggregation Constraints", mimeo

Qi, M. (2001), "Predicting US recessions with leading indicators via neural network models", *International Journal of Forecasting* 17: 383-401.

Raj, B. (2002), “Asymmetries of Business Cycles: the Markov-Switching Approach”, in A. Ullah, A. Wan and A. Chaturvedi (eds.), Handbook of Applied Econometrics and Statistical Inference, Marcel Dekker, Ch. 31, 687-710

Ravn, M. O. and M. Sola (1999), "Business cycle dynamics: predicting transitions with macrovariables", in: P. Rothman, ed., *Nonlinear time series analysis of economic and financial data*, vol. 1 (Kluwer Academic Publishers), ch. 12: 231-265.

Sargent, T. J. and C. A. Sims (1977), "Business cycle modeling without pretending to have too much a priori economic theory", in: C. Sims et al., *New methods in business cycle research*, Federal Reserve Bank of Minneapolis.

Shepard, N. (1994), "Partial non-gaussian state space", *Biometrika* 81: 115-131.

Sichel, D.E. (1994), "Inventories and the Three Phases of the Business Cycle," *Journal of Business and Economic Statistics*, vol. 12 (July), pp. 269-77.

Simpson, P. W., Osborn, D. R., and M. Sensier (1999), "Modelling business cycle movements in the UK economy", School of Economic Studies Discussion Paper Series, University of Manchester, no. 9908.

Sims, C. A. (1989), "Comment on Stock and Watson (1989)", NBER Macroeconomics Annual: 394-39

Sims, C. A., Stock, J. H. and M. W. Watson (1990), "Inference in linear time series models with some unit roots", *Econometrica* 58: 113-144.

Stock, J. H. and M. W. Watson (1989), "New indexes of coincident and leading economic indicators", in: Blanchard, O., and S. Fischer, eds., NBER Macroeconomics Annual, MIT Press (Cambridge, MA): 351-394.

Stock, J. H. and M. W. Watson (1991), "A probability model of the coicident indicators", in Lahiri, K., and G. H. Moore, eds., *Leading Economic Indicators: New approaches and forecasting records*, Cambridge University Press (Cambridge, UK).

Stock, J. H. and M. W. Watson (1992), "A procedure for predicting recessions with leading indicators: econometric issues and recent experience", NBER Working Paper Series, no. 4014.

Stock, J. H. and M. W. Watson (1993), "A procedure for predicting recessions with leading indicators: econometric issues and recent experience", in: Stock, J. H., and M. W. Watson, eds., *Business Cycles, Indicators, and Forecasting,* The University of Chicago Press (Chicago): 95-153.

Stock, J. H. and M. W. Watson (1999a), "Business cycle fluctuations in US macroeconomic time series", in: Taylor, J. B., and M. Woodford, eds., *Handbook of Macroeconomics*, vol. IA, North-Holland (Amsterdam).

Stock, J. H. and M. W. Watson (1999b), " A Comparison of Linear and Nonlinear Univariate Models for Forecasting Macroeconomic Time Series", in: Cointegration, Causality, and Forecasting - Festschrift in Honour of Clive W. J. Granger, edited by R. Engle and H. White

Stock, J. H. and M. W. Watson (2002a), "Macroeconomic Forecasting Using Diffusion Indexes", *Journal of Business and Economic Statistics*, 20, 147-62.

Stock, J. H. and M. W. Watson (2002b), "Forecasting Using Principal Components from a Large Number of Predictors", *Journal of the American Statistical Association*, 97, 1167--1179.

Stock, J. H. and M. W. Watson (2003a), "Forecasting output and inflation: the role of asset prices", NBER Working Paper no. 8180.

Stock, J. H. and M. W. Watson (2003b), "How did the leading indicator forecasts perform during the 2001 recession", Federal Reserve Bank of Richmond Economic Quarterly 89: 71-90

Stock, J. H. and M. W. Watson (2004), "Forecasting with Large Datasets", in G. Elliott, C.W.J. Granger and A. Timmermann (eds.) Handbook of Economic Forecasting

Swanson, N. R., Ghysels, E. and M. Callan (1998), "A multivariate time series analysis of the data revision process for industrial production and the composite leading indicator", Working Paper (Pennsylvania State University).

Swanson, N. R. and H. White (1997), "A model selection approach to real-time macroeconomic forecasting using linear models and artificial neural networks", *The Review of Economics and Statistics* 79: 540-550.

Terasvirta, T. (2004), "Forecasting with Nonlinear Models", in G. Elliott, C.W.J. Granger and A. Timmermann (eds.) Handbook of Economic Forecasting

Timmermann, A.G. (2004) "Forecast Combination", in G. Elliott, C.W.J. Granger and A. Timmermann (eds.) Handbook of Economic Forecasting

Vaccara, B. N., and V. Zarnowitz (1997), "How good are the leading indicators?", Proceedings of the Business and Economic Statistics Section (American Statistical Association).

Walsh, C.E. (2003), Monetary Theory and Policy, 2nd. ed., The MIT Press.

Watson, M. (1991), "Using Econometric Models to Predict Recessions" *Economic Perspectives*, (Research Periodical of the Chicago Federal Reserve Bank), September/October

Wecker, W. E. (1979), "Predicting the turning points of a time series", *Journal of business* 52(1): 35-50.

West, K. D. (1996), "Asymptotic inference about predictive ability", *Econometrica* 64: 1067-1084.

West, K. (2004), "Forecast Evaluation", in G. Elliott, C.W.J. Granger and A. Timmermann (eds.), Handbook of Economic Forecasting

Zarnowitz, V. and C. Boschan (1975a), "Cyclical indicators: an evaluation and new leading indexes", *Business Condition Digest*, May 1975, reprinted in: *Handbook of Cyclical Indicators*, 1977: 170-183.

Zarnowitz, V. and C. Boschan (1975b), "New composite indexes of coincident and lagging indicators", *Business Condition Digest*, November 1975, reprinted in: *Handbook of Cyclical Indicators*, 1977: 185-198.

Zarnowitz, V. and P. Braun (1989), "Major macroeconomic variables and leading indicators: some estimates of their interrelations, 1886-1982", NBER Working Paper no. 2812.

Zellner, A., C. Hong and G.M. Gulati (1990), "Turning Points in Economic Time Series, Loss Structures and Bayesian Forecasting" in S. Geisser, J. Hodges, S.J. Press, and A. Zellner (eds.), *Bayesian and Likelihood Methods in Statistics and Econometrics: Essays in Honor of George A. Barnard*, Amsterdam: North-Holland, 371-393.

Zellner, A. and C.Hong (1991), "Bayesian Methods for Forecasting Turning Points in Economic Time Series: Sensitivity of Forecasts to Asymmetry of Loss Structures" in K. Lahiri and G.H. Moore (eds.), *Leading Economic Indicators: New Approaches and Forecasting Records*, England: Cambridge Univ. Press, 129-140.

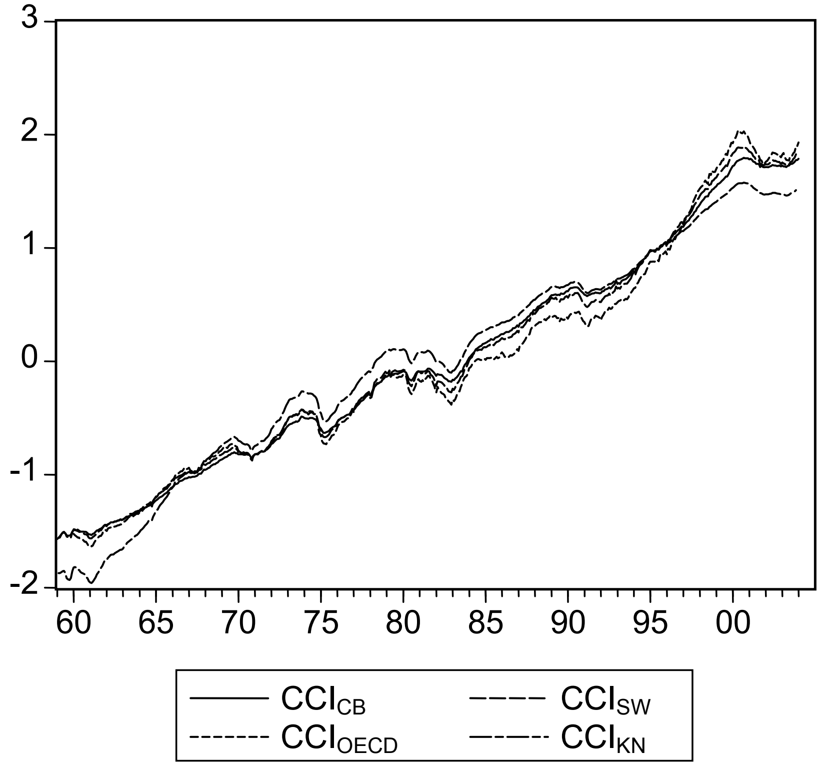
Zellner, A., C. Hong and C. Min (1991), "Forecasting Turning Points in International Output Growth Rates Using Bayesian Exponentially Weighted Autoregression, Time-Varying Parameter and Pooling Techniques", *Journal of Econometrics*, 49, 275-304.

Zellner, A. and C. Min (1994), “Bayesian and Non-Bayesian Methods for Combining Models and Forecasts with Applications to Forecasting International Growth Rates", Journal of Econometrics, 56, 89-118.

Zellner, A. and P.E. Rossi (1984), "Bayesian Analysis of Dichotomous Quantal Response Models", *Journal of Econometrics* 25 (1984), 365-394.

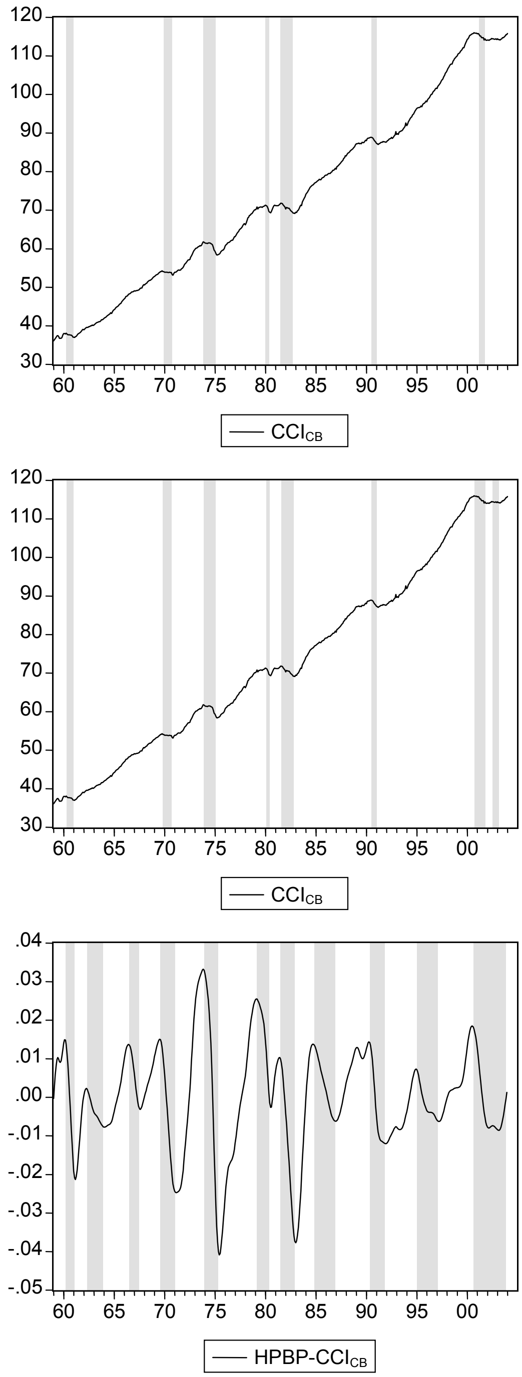
Zha, T. (1998), "A Dynamic Multivariate Model for Use in Formulating Policy", *Economic Review* 83 (First Quarter 1998).

Figure 1: Composite Coincident Indexes



Note: The figure reports the Conference Board's composite coincident indicator (CCI_{CB}), the OECD reference coincident series (CCI_{OECD}), Stock and Watson's coincident index (CCI_{SW}), and the coincident index derived from the four components in CCI_{CB} modeled with a dynamic factor model as in Kim and Nelson (1998) (CCI_{KN}). All indexes have been normalized to have zero mean and unit standard deviation.

Figure 2: Classical and deviation cycles

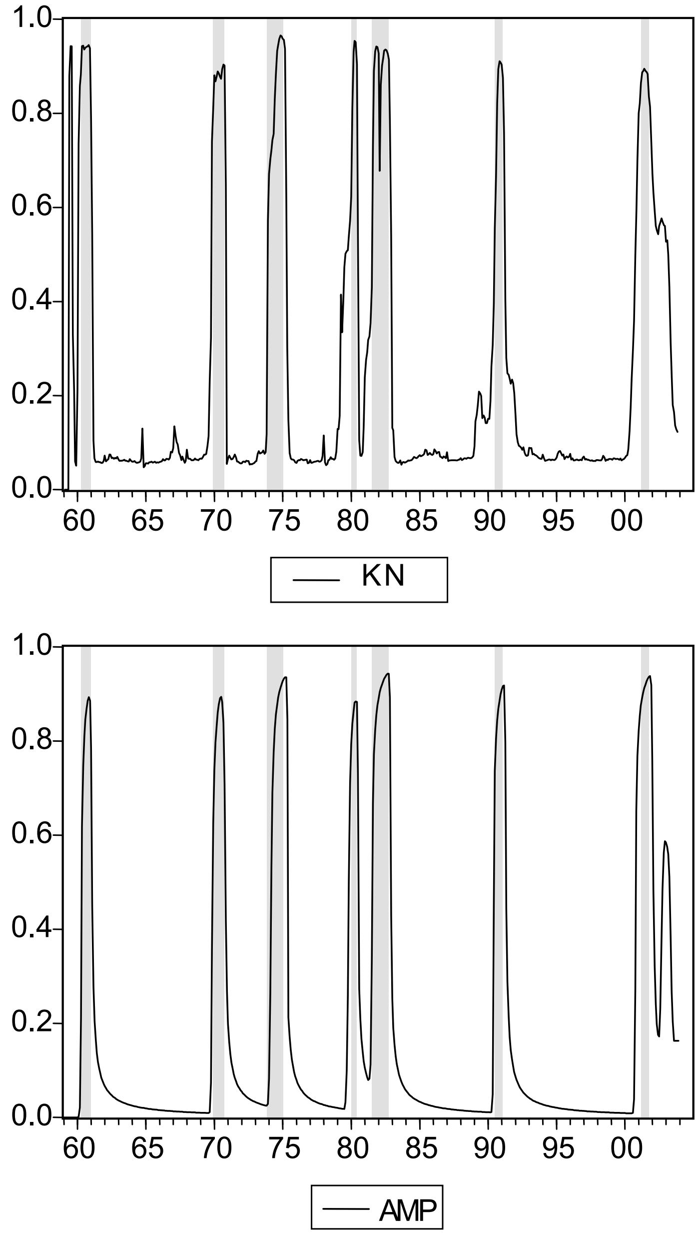


Note: Upper panel: CCI_{CB} and NBER dated recessions (shaded areas).

Middle panel: CCI_{CB} and recessions dated with Artis, Marcellino, Proietti (2003) algorithm (shaded areas).

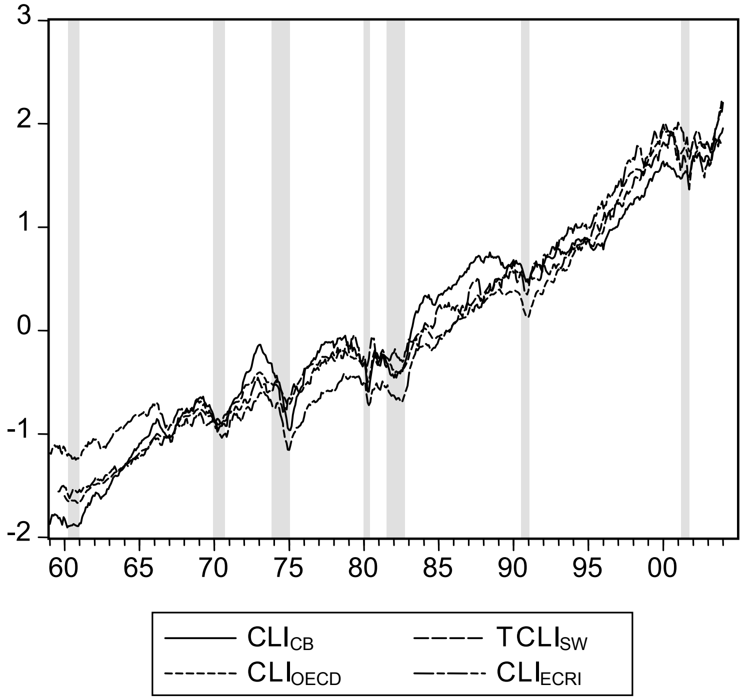
Lower panel: HP-band pass filtered CCI_{CB} and recessions dated with Artis, Marcellino, Proietti (2003) algorithm (shaded areas).

Figure 3: Probability of recession and NBER dated recessions



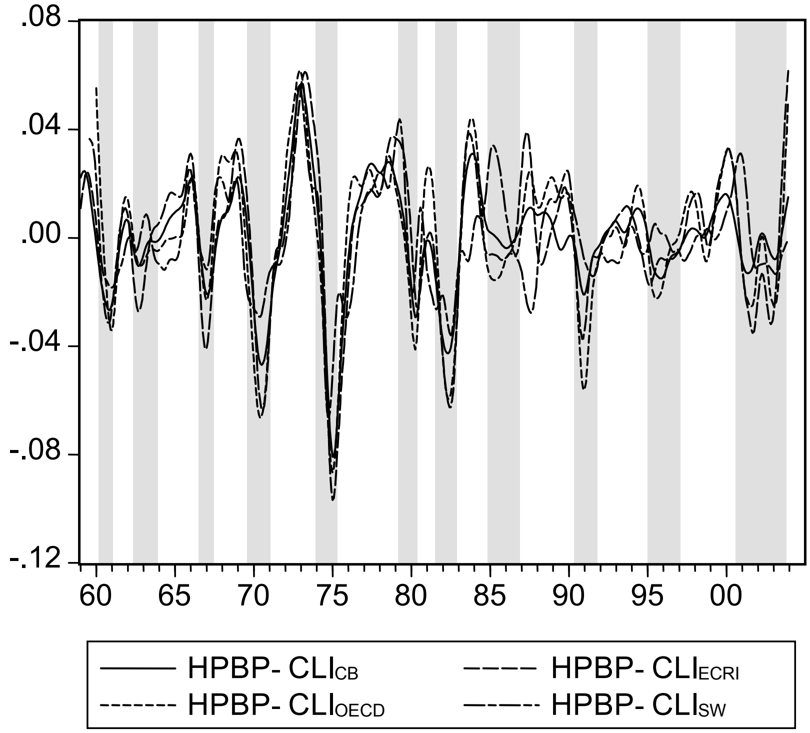
Note: The upper panel reports the (filtered) probability of recession computed from a dynamic factor model for the four components in the CCI_{CB} using the Kim and Nelson's (1998) methodology.
The lower panel reports the (filtered) probability of recession computed using the algorithm in Artis, Marcellino, Proietti (2003) applied to the CCI_{CB}.
The shaded areas are the NBER dated recessions.

Figure 4: Composite Leading Indexes



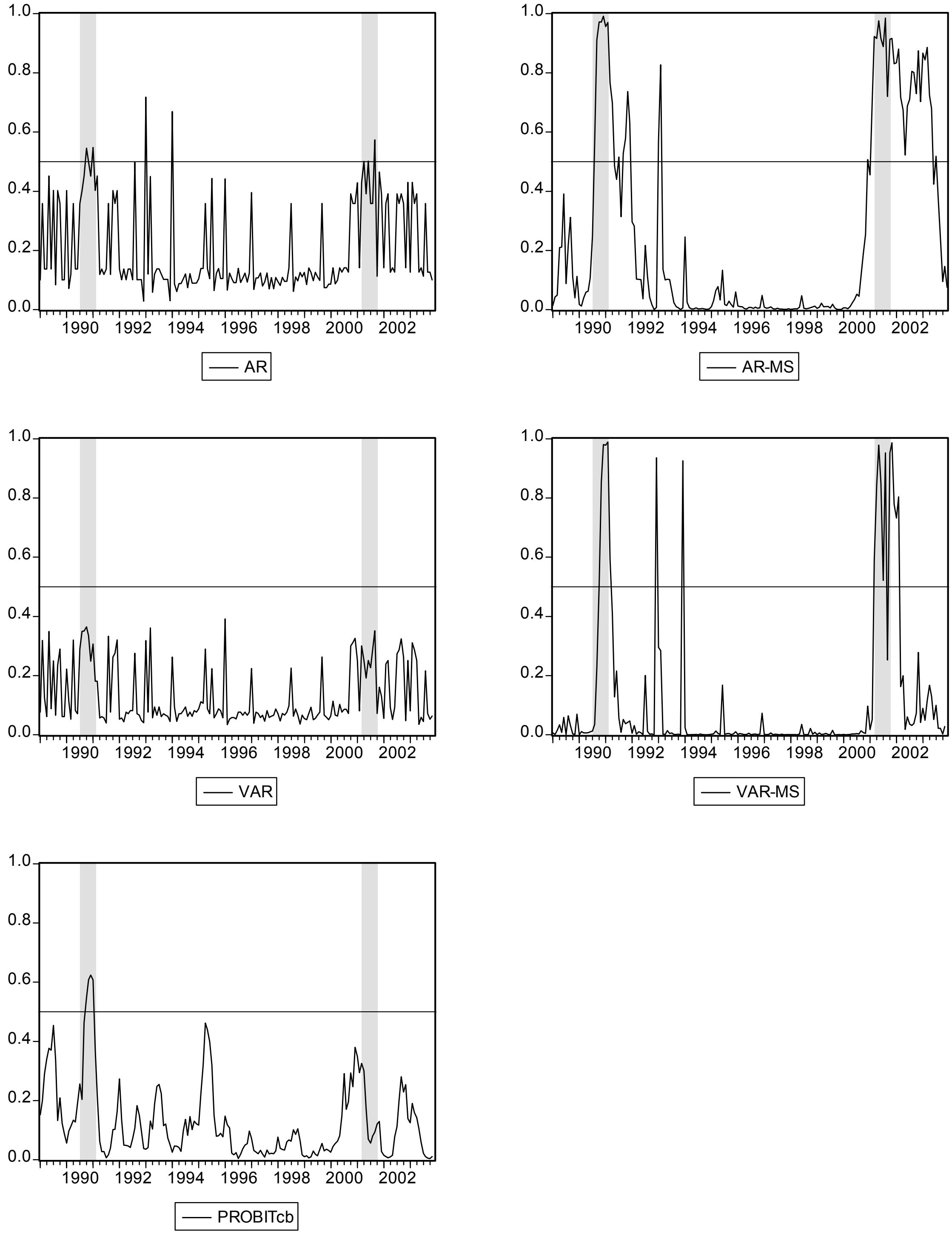
Note: The figure reports the Conference Board composite leading index (CLI_{CB}), the OECD leading index (CLI_{OECD}), a transformation of Stock and Watson's leading index ($TCLI_{SW}$, see text), the ECRI leading index (CLI_{ECRI}), and the NBER dated recessions (shaded areas). All indexes have been normalized to have zero mean and unit standard deviation.

Figure 5: Filtered composite leading indexes with AMP dated recessions for deviation cycle of CCI_{CB}



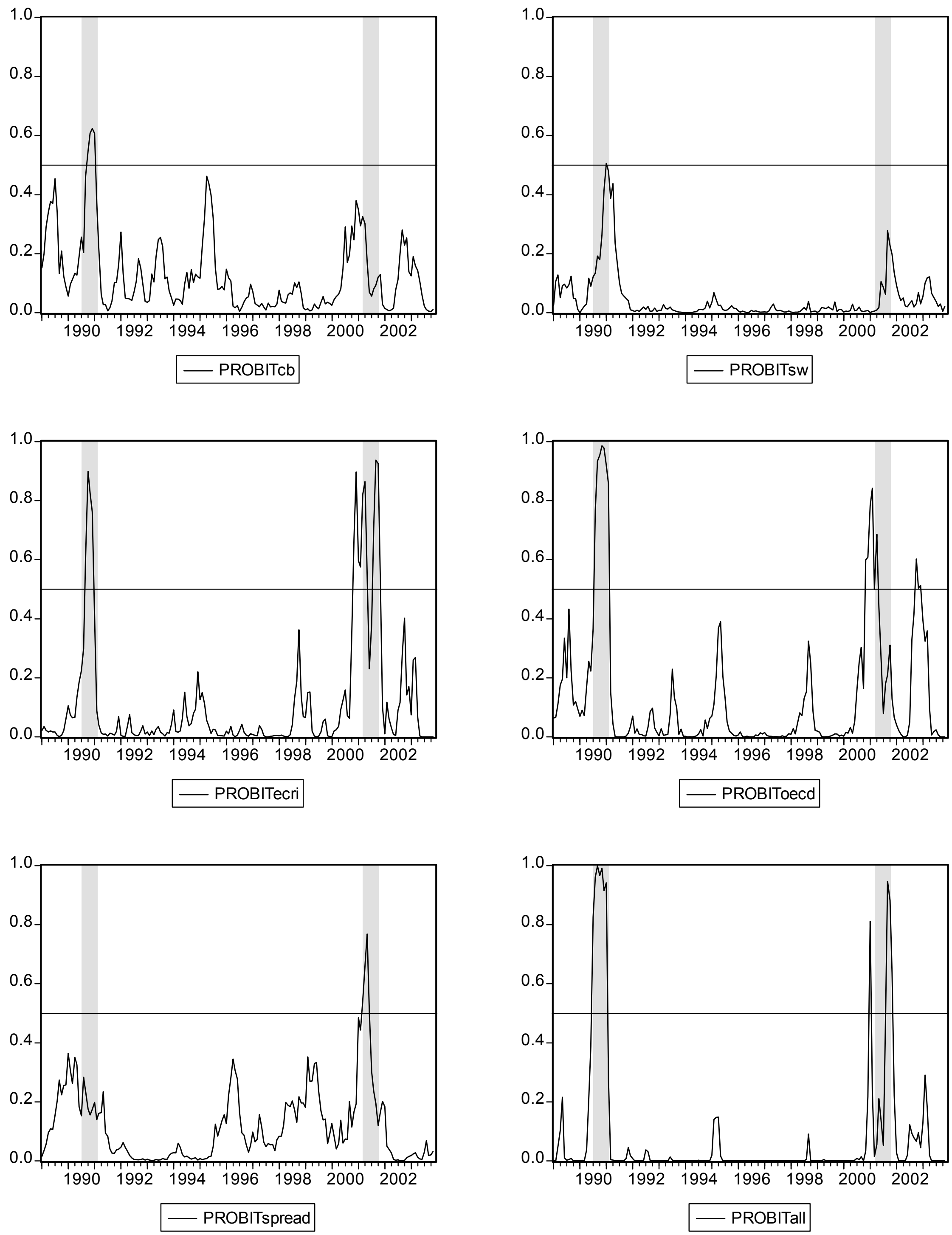
Note: The figure reports the HP-band pass filtered versions of the four CLIs in Figure 4, and the Artis, Marcellino, Proietti (2003) dating of the HP band pass filtered versions of the CCI_{CB} (shaded areas).

Figure 6: One month ahead recession probabilities



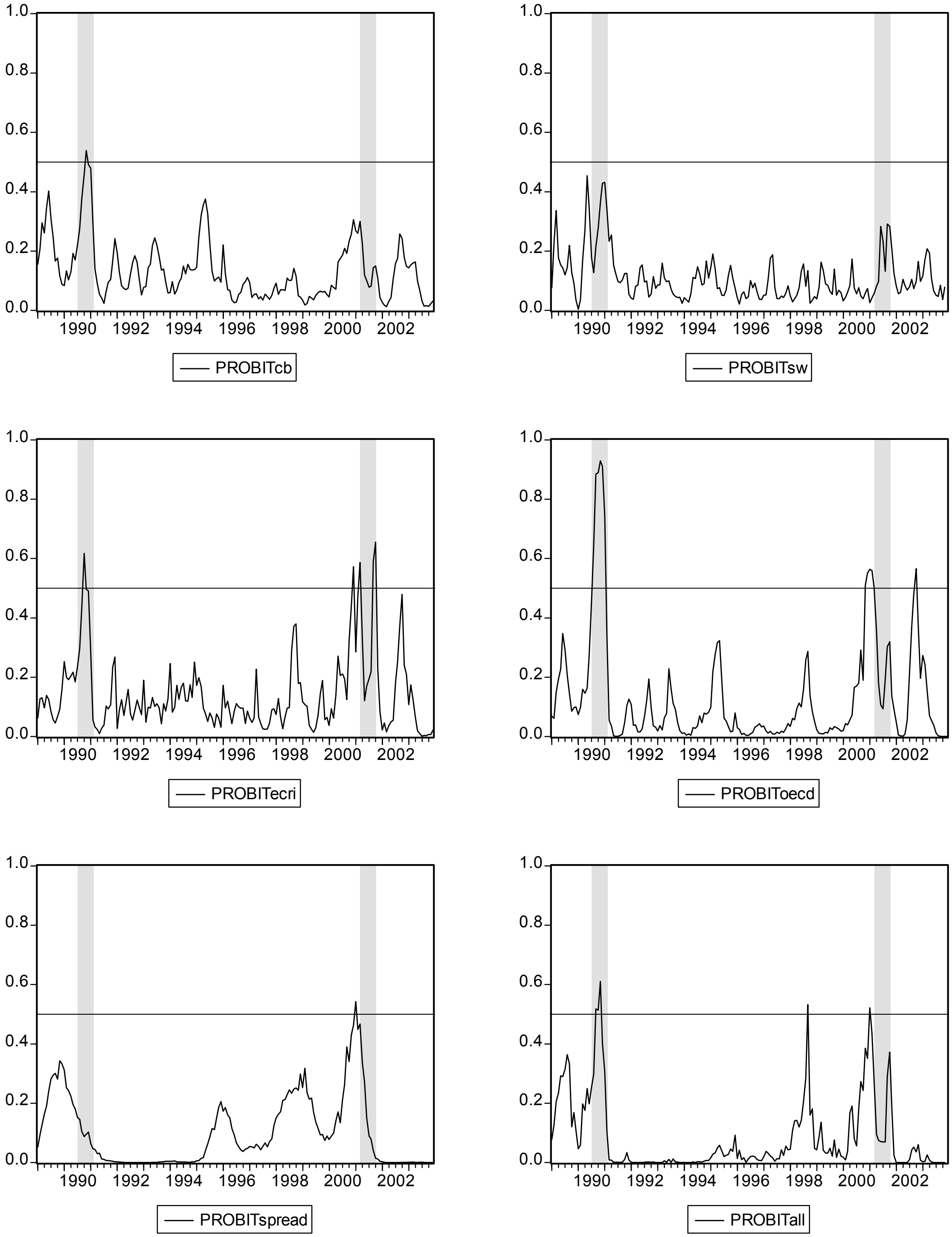
Note: The models are those in Table 7. Shaded areas are NBER dated recessions.

Figure 7: One month ahead recession probabilities for alternative probit models



Note: The models are those in Table 8. Shaded areas are NBER dated recessions.

Figure 8: Six months ahead recession probabilities for alternative probit models



Note: The models are those in Table 8. Shaded areas are NBER dated recessions.

Table 1: Correlation of composite coincident indexes (6-month percentage change)

	CCI_{CB}	CCI_{OECD}	CCI_{SW}	CCI_{KN}
CCI_{CB}	1			
CCI_{OECD}	0.941	1		
CCI_{SW}	0.979	0.969	1	
CCI_{KN}	0.943	0.916	0.947	1

Note: Common sample is 1970:01 – 2003:11.

Table 2: Correlation of composite leading indexes (6-month percentage change)

	CLI_{CB}	CLI_{OECD}	CLI_{sw}	CLI_{ECRI}
CLI_{CB}	1			
CLI_{OECD}	0.891	1		
CLI_{sw}	0.719	0.601	1	
CLI_{ECRI}	0.817	0.791	0.595	1

Note: Common sample is 1970:01 – 2003:11.

Table 3: Classical cycles, dating of coincident and leading indexes

Peak						Trough					
Coincident		Leading (AMP)				Coincident		Leading (AMP)			
NBER	AMP	CB	OECD	ECRI	SW	NBER	AMP	CB	OECD	ECRI	SW
Apr 1960	May 1960	Jan 1959 *	Jan 1960 *	Jan 1959	Aug 1959 *	Feb 1961	Feb 1961	Mar 1960	Dec 1960	Oct 1960	May 1960
				Jan 1962						Jun 1962	
		Apr 1966	Apr 1966	Apr 1966	Feb 1966			Dec 1966	Nov 1966	Dec 1966	Jul 1966
Dec 1969	Nov 1969	May 1969	Jan 1969	Jan 1969	**MISSING**	Nov 1970	Nov 1970	Apr 1970	Apr 1970	Jul 1970	**MISSING**
Nov 1973	Dec 1973	Feb 1973	Feb 1973	Jun 1973	Jan 1973	Mar 1975	Mar 1975	Jan 1975	Dec 1974	Jan 1975	Aug 1974
Jan 1980	Feb 1980	Nov 1978	Aug 1978	Nov 1978	Jun 1979	Jul 1980	Jul 1980	Apr 1980	Apr 1980	May 1980	Aug 1981
Jul 1981	Aug 1981	Nov 1980	Nov 1980	May 1981	**MISSING**	Nov 1982	Dec 1982	Jan 1982	Feb 1982	Aug 1982	**MISSING**
			Feb 1984		Oct 1985				Sep 1984		Jun 1986
		Jul 1988						Jun 1989			
Jul 1990	Jul 1990	Feb 1990	Mar 1990	Oct 1989	Feb 1990	Mar 1991	Mar 1991	Jan 1991	Dec 1990	Dec 1990	Jan 1991
		Nov 1994	Dec 1994					May 1995	Apr 1995		
				May 1998						Oct 1998	
Mar 2001	Oct 2000	Feb 2000	Feb 2000	Feb 2000	**MISSING**	Nov 2001	Dec 2001	Mar 2001	Oct 2001	Oct 2001	**MISSING**
	Jul 2002	**MISSING**	May 2002	**MISSING**	Feb 2002		Apr 2003	**MISSING**	**MISSING**	Apr 2003	**MISSING**

	Peak CB		Peak OECD		Peak ECRI		Peak SW		Trough CB		Trough OECD		Trough ECRI		Trough SW	
	NBER	AMP	NBER	AMP	NBER	AMP	NBER	AMP	NBER	AMP	NBER	AMP	NBER	AMP	NBER	AMP
Average Lead	10	11	9	9	9	10	7	8	9	9	4	4	3	3	8	9
St. Dev.	4.23	4.28	4.30	5.31	5.13	4.75	3.78	2.50	4.30	5.31	2.89	3.04	1.11	1	5.38	5.80
False Alarms	3	3	3	3	3	3	2	2	3	3	3	3	3	3	2	2
Missing	0	1	0	0	0	1	2	4	0	1	0	0	0	1	3	4

Note: Shaded values are false alarms, 'MISSING' indicates a missed turning point. Leads longer than 18 months are considered false alarms. Negative leads are considered missed turning points. * indicates no previous available observation. Based on final release of data.
AMP: Dating based on algorithm in Artis, Marcellino, Proietti (2003).

Table 4: Correlations of HP band pass filtered composite leading indexes

	HPBP-CLI_{CB}	HPBP-CLI_{OECD}	HPBP-CLI_{ECRI}	HPBP-CLI_{SW}
HPBP-CLI_{CB}	1			
HPBP-CLI_{OECD}	0.919	1		
HPBP-CLI_{ECRI}	0.906	0.882	1	
HPBP-CLI_{SW}	0.703	0.595	0.645	1

Note: Common sample is 1970:01 – 2003:11.

Table 5: Deviations cycles, dating of coincident and leading indexes

Peak					Trough				
Coincident	Leading				Coincident	Leading			
CB	CB	OECD	ECRI	SW	CB	CB	OECD	ECRI	SW
Mar 1960	May 1959	Feb 1960	Jul 1959	Sep 1959	Mar 1961	Nov 1960	Jan 1961	Oct 1960	Jan 1961
May 1962	Jan 1962	Jan 1962	Dec 1961	**MISSING**	Jan 1964	Sep 1962	Nov 1962	Sep 1962	**MISSING**
				Apr 1963					May 1964
Jul 1967	Feb 1966	Mar 1966	Feb 1966	Jan 1967	Aug 1967	Feb 1967	Jan 1967	Dec 1966	Jan 1967
Aug 1969	Feb 1969	Dec 1968	Feb 1969	Dec 1967	Mar 1971	Jul 1970	Jun 1970	Aug 1970	Jun 1970
Dec 1973	Feb 1973	Jan 1973	May 1973	Jan 1973	Jun 1975	Feb 1975	Jan 1975	Jan 1975	Oct 1974
Mar 1979	Sep 1978	Sep 1978	Dec 1978	May 1979	Jul 1980	May 1982	Apr 1980	Jun 1982	Feb 1980
Jul 1981	**MISSING**	Mar 1981	**MISSING**	Sep 1980	Jan 1983	**MISSING**	May 1982	**MISSING**	Jun 1982
Nov 1984	Jan 1984	Dec 1983	Oct 1983	Apr 1985	Jan 1987	Jan 1986	May 1985	Oct 1985	Aug 1987
			Jun 1987					Apr 1988	
May 1990	Sep 1987	Aug 1987	Nov 1989	Jan 1990	Dec 1991	Dec 1990	Jan 1991	Nov 1990	Jul 1991
		Feb 1993					Jul 1993		
Jan 1995	Jun 1994	Jun 1994	Oct 1993	Jan 1994	Mar 1997	Nov 1995	Aug 1995	Feb 1995	Oct 1994
				Aug 1995					May 1997
		Nov 1997					Oct 1998		
Aug 2000	Jan 2000	Mar 2000	Mar 2000	Jan 2001	Dec 2003 *	May 2001	Dec 2003 *	Dec 2003 *	Nov 2003 *
	May 2002					Dec 2003 *			
Aver. Lead	7	6	7	8		10	7	10	6
St. Dev.	2.28	3.21	3.80	3.25		4.67	4.03	4.47	2.31
False Alarms	2	2	1	2		1	4	2	1
Missing	1	0	1	4		1	0	1	3

Note: Shaded values are false alarms, 'MISSING' indicates a missed turning point. Leads longer than 18 months are considered false alarms. Negative leads are considered missed turning points. * indicates last available observation. Based on final release of data. AMP: Dating based on algorithm in Artis, Marcellino, Proietti (2003).

Table 6: Forecast comparison of alternative VAR models for CCI_{CB} and CLI_{CB}

		1 step-ahead		6 step-ahead DYNAMIC		6 step-ahead ITERATED	
		Relative MSE	Relative MAE	Relative MSE	Relative MAE	Relative MSE	Relative MAE
		whole sample					
CCI + CLI	VAR(2)	1	1	1	1	1	1
CCI	AR(2)	1.001	1.010	0.982	0.963 *	1.063	1.032
CCI + CLI coint	VECM(2)	1.042	1.074 *	1.067	1.052	1.115	1.100
4 comp. of CCI + CLI	VAR(2)	0.904 **	0.976	0.975	0.973	0.854 **	0.911 **
CCI + 10 comp. of CLI	VAR(1)	1.158 ***	1.114 ***	1.035	1.017	1.133 **	1.100 ***
4 comp. CCI + 10 comp. CLI	VAR(1)	0.995	1.029	1.090	1.035	0.913	0.967
		MSE	MAE	MSE	MAE	MSE	MAE
	VAR(2)	0.075	0.186	0.079	0.216	0.075	0.201
		recessions					
CCI + CLI	VAR(2)	1	1	1	1	1	1
CCI	AR(2)	0.988	0.975	0.949	0.940	1.303 **	1.154 **
CCI + CLI coint	VECM(2)	0.681 ***	0.774 ***	0.744	0.882	0.478 ***	0.626 ***
4 comp. of CCI + CLI	VAR(2)	0.703 *	0.784 **	0.825	0.879	0.504 ***	0.672 ***
CCI + 10 comp. of CLI	VAR(1)	1.095	1.009	1.151	1.131	1.274 *	1.117
4 comp. CCI + 10 comp. CLI	VAR(1)	0.947	0.852	1.037	1.034	0.614 ***	0.714 ***
		MSE	MAE	MSE	MAE	MSE	MAE
	VAR(2)	0.087	0.258	0.096	0.252	0.163	0.368
		expansions					
CCI + CLI	VAR(2)	1	1	1	1	1	1
CCI	AR(2)	1.002	1.016	0.977	0.956 *	0.997	1.005
CCI + CLI coint	VECM(2)	1.090 *	1.123 ***	1.118	1.081	1.292 ***	1.206 ***
4 comp. of CCI + CLI	VAR(2)	0.931 *	1.007	0.987	0.980	0.952	0.964
CCI + 10 comp. of CLI	VAR(1)	1.166 ***	1.132 ***	1.015	0.997	1.093 *	1.096 **
4 comp. CCI + 10 comp. CLI	VAR(1)	1.001	1.058	1.087	1.029	0.997	1.023
		MSE	MAE	MSE	MAE	MSE	MAE
	VAR(2)	0.074	0.177	0.076	0.208	0.065	0.183

Note: Forecast sample is: 1989:1 – 2003:12. First estimation sample is 1959:1 – 1988:12 (for 1 step-ahead) or 1959:1 – 1988:6 (for 6 step-ahead), recursively updated. Lag lenght selection by BIC. MSE and MAE are mean square and absolute forecast error. VAR for CCI_{CB} and CLI_{CB} is benchmark. *, **, *** indicate significance at 10%, 5%, 1% of the Diebold-Mariano test for the null hypothesis of no significant difference in MSE or MAE with respect to the benchmark.

Table 7: Turning point predictions

Target	Model	Relative MSE	Relative MAE	
NBER (1 step-ahead)	univariate	1.0302	1.2685	***
	univariate MS	1.3417	1.0431	
	bivariate	1.0020	1.0512	
	bivariate MS	0.6095	0.4800	***
	probit CLI_CB	1	1	
		MSE	MAE	
	probit	0.0754	0.1711	

Note: One-step ahead turning point forecasts for the NBER expansion/recession indicator. Linear and MS models (as in Hamilton and Perez-Quiros (1996)) for CCI_{CB} and CLI_{CB}. Six lags of CLI_{CB} are used in the probit model. *, **, *** indicate significance at 10%, 5%, 1% of the Diebold-Mariano test for the null hypothesis of no significant difference in MSE or MAE with respect to the benchmark.

Table 8: Forecasting performance of alternative CLIs using probit models for NBER recession/expansion classification

Target	Model	Relative MSE		Relative MAE	
NBER (1 step-ahead)	CLI_CB	1		1	
	CLI_SW	1.01		0.664	***
	CLI_ECRI	0.588		0.597	***
	CLI_OECD	0.719		0.714	***
	termspread	0.952		0.937	
	4 CLI+spread	0.565	**	0.404	***
NBER (6 step-ahead)	CLI_CB	1		1	
	CLI_SW	1.085		0.956	
	CLI_ECRI	0.888		0.948	
	CLI_OECD	0.912		0.834	**
	termspread	0.736	**	0.726	***
	4 CLI+spread	0.837	**	0.692	***
		MSE		MAE	
CLI_CB	1 step-ahead	0.073		0.169	
	6 step-ahead	0.085		0.191	

Note: Forecast sample is: 1989:1 – 2003:12. First estimation sample is 1959:1 – 1988:12, recursively updated. Fixed lag lenght: 6 lags for the first four models and 3 lags for the model with all four CLIs (see text for details). MSE and MAE are mean square and absolute forecast error. Probit model for CLI_{CB} is benchmark. *, **, *** indicate significance at 10%, 5%, 1% of the Diebold-Mariano test for the null hypothesis of no significant difference in MSE or MAE with respect to the benchmark.

Table 9: Evaluation of forecast pooling

Combine	Relative MSE		Relative MAE		Relative MSE		Relative MAE	
	Predicting CCI_CB growth							
	MSE-weighted				simple average			
6 linear models (1month)	0.9474	**	0.9824		0.9418	**	0.9781	
6 linear models (6month dynamic)	0.8873		0.9100		0.8863		0.9082	
6 linear models (6month iterated)	0.9352	**	0.9776		0.9255	**	0.9701	
	Predicting NBER turning points							
	MSE-weighted				simple average			
4 linear and MS models (1m)	0.8683		1.1512		0.6676		0.9607	
4 linear and MS models + probit (1m)	0.8300		1.0989		0.6695		0.9686	
	Predicting NBER turning points							
	MSE-weighted				simple average			
5 single index PROBIT (1m)	0.7423	**	0.8028	***	0.7014	**	0.7844	***
5 single index PROBIT + all (1m)	0.6900	**	0.7579	***	0.6395	**	0.7234	***
5 single index PROBIT (6m)	0.8863	***	0.9069	**	0.8667	***	0.8956	**
5 single index PROBIT + all (6m)	0.8707	***	0.8695	***	0.8538	***	0.8569	***

Note: Forecast sample is 1989:1 – 2003:12. The forecasts pooled in the upper panel are from the six models in Table 6 and the benchmark is the VAR(2). The forecasts pooled in the middle panel are from the models in Table 7, including or excluding the probit, and the benchmark is the probit model with 6 lags of CLI_{CB} as regressor. The forecasts pooled in the lower panel are from the models in Table 8, including or excluding the probit with all indicators, and the benchmark is as in the middle panel. *, **, *** indicate significance at 10%, 5%, 1% of the Diebold-Mariano test for the null hypothesis of no significant difference in MSE or MAE with respect to the benchmark.