# Detecting Turning Points with Many Predictors through Hidden Markov Models

**Benoît BELLONE[A], David SAINT-MARTIN[B]**

**Working Paper**

**(First draft :December 2003)**

*The views expressed in this working paper are those of the authors and do not necessarily represent those of the French Forecasting and Economic Analysis Directorate or the Ecole Nationale des Ponts et Chaussées. This working paper describes research in progress by the authors and is published to elicit comments and to further debate.*

[A] French Forecasting and Economic Analysis Directorate, e-mail: benoit.bellone@dp.finances.gouv.fr
[B] Ecole Nationale des Ponts et Chaussées, e-mail: david.saint-martin@polytechnique.org

This draft: December 2003[1]

**Résumé:**

Cet article étudie le "Cycle d'Affaire" Américain à l'aide de modèles à changements de régimes markoviens (HMM) appliqués à des données mensuelles. Il permet d'extraire dix séries particulièrement robustes pour détecter des phases de ralentissement en temps réel. Ce papier présente différents modèles de détection de récession basés sur des processus markoviens et examine leur performance. Il expertise notamment la méthode la plus simple et la plus efficace pour synthétiser l'information à travers ces modèles. On tire de ces travaux trois conclusions principales: de simples modèles HMM sont déterminants pour suivre le cycle d'affaire à travers quelques séries bien choisies et particulièrement fiables; des modèles plus complexes tels que le modèle à facteur dynamique avec changement de régime (DFMS) ou l'indicateur coïncident probabiliste de Stock et Watson ne paraissent pas plus puissants que de simples modèles à changement de régimes markovien (univariés ou pseudo-multivariés); aggréger l'information dans l'espace des données semble conduire à de meilleures performance que la construction d'un résumé en combinant les probabilités pour des données " haute-fréquence". Ce papier conclue sur les propriétés de ces modèles en terme d'indicateurs avancés et de détection sur données en temps réel et donne quelques pistes de recherche complémentaire.

**Abstract:**

This paper explores the American business cycle with the Hidden Markov Model (HMM) as a monitoring tool using monthly data. It exhibits ten US time series which offer reliable information to detect recessions in real time. It also proposes and assesses the performances of different and complementary "recession models" based on Markovian processes, discusses the most efficient and easiest way of encompassing information through these models and draws three main conclusions: simple HMM are decisive to monitor the business cycle and some series are proved highly reliable; more sophisticated models such as the Dynamic Factor with Markov Switching (DFMS) model or Stock and Watson's Experimental Recession Index seem not to be more powerful than simple (univariate or pseudo-multivariate) Hidden Markov Models, which remain far more parsimonious; combining information in temporal space seems to work marginally better than in probability space for high frequency data. We conclude about leading and "real time detection" properties related to HMM and give some hints for further research.

**KEYWORDS**: Business Cycle, Markov Switching, Dynamic Factor, Coincident Indicators

**JEL Classification** : C32, E32, E44

[1] We are grateful to Vladimir Borgy for his material help during all the stages of this research, to Eric Dubois for his support. At the start of this research, David Saint-Martin was a graduate student at the Ecole Polytechnique and Benoit Bellone an economist at the Forecasting Division of the French Forecasting and Economic Analysis Directorate. A preliminary version of this working paper was presented in Paris on December 3rd 2003 at a "Seminaire Fourgeaud" dedicated to "Short term Forecasting". A special thank to Franck Sédillot and Fabrice Lenglart for their remarks and advice. The views expressed in this working paper are those of the authors and do not necessarily represent those of the French Forecasting and Economic Analysis Directorate, of the Ecole Polytechnique or the Ecole Nationale des Ponts et Chaussées.

# Contents

# 1. Introduction

Much attention has already been drawn to measure the American business cycle and detect its turning points. However, as reminded by Stock (2003), the last US recession caught the forecasters unawares: "*Even as late as the fourth quarter of 2000, when industrial production was already declining, the median Survey of Professional Forecasters(SPF) conducted by the Federal Reserve Bank of Philadelphia was predicting strong economic growth throughout 2001.[…] Throughout the first three quarters of 2001*, [the current quarters of recession defined by the NBER, the probability of a running recession] *hovered around one third among the SPF. When in the fourth quarter of 2001[…] the SPF forecasters finally were sure that growth would be negative, the consensus probability of negative growth was 82%, the economy was indeed running out of recession*". The NBER dating release was also of little help to assess the peak and trough in real time[2]. Could some coincident probabilistic indicators have beaten this poor performance?

From the early works of Burns and Mitchell (1946) who developed the concepts of co-movement and asymmetric shifts in expansion or contraction periods, to the decisive papers of Hamilton (1989), Stock and Watson (1989) and more recently Kim and Nelson (1998) much headway has been made in detecting cycles through stochastic modeling. Following the general theory on statistical surveillance as described in Frisen and de Marè (1991), there are proofs for the optimal properties of methods based on likelihood or posterior distributions. Their extension to a multivariate framework suggested by Diebold and Rudebusch (1996) gave rise to a rich literature based on the Dynamic Factor with Markov Switching (DFMS) model as presented in Kim and Yoo (1995) and Chauvet (1998). But all these models only focused on the four indicators of Business Coincident Index (BCI) released by the Conference Board: Industrial Production, Real manufacturing and trade sales, Real personal income and Total Employee-hours in non agricultural establishment. Although the four indicators described above are the most important measures considered by the NBER in developing its business cycle chronology, there is no fixed rule about which other measures contribute information to the process. As a result, these four series do not contain all the necessary knowledge to characterize business cycle dynamics in real time, not to mention that they are constantly revised. Thus, there should be room for identifying other potent indicators. However as emphasized by Watson (2001), the choice of the variables and their number remains crucial: as many time series may have marginal predictive content there may be large gains from increasing the number of predictors from 1 to 10, but negligible ones from increasing it beyond.

The novel aspect of this paper is twice. First, through discussing the use of Hidden Markov Models, it aims at identifying new reliable time series preferably exempt of revisions to challenge the Business Coincident Index (BCI). Second, it tries to solve the puzzle of "how to combine information?". Thus, we tried to assess the marginal gain of "pooled data models", that is following the insights of Watson (2001) that "averaging" is a simple, but apparently effective model forecasting approach among the family of "factor models". The main results of our research can then be summarized:

HMM are decisive to monitor the business cycle and even growth cycle, providing that we use different rules to extract information. In fact, as suggested by Chauvet (1998), even coincident series can send early warnings if filtered probabilities are carefully monitored. Besides, these models carry evidences of high stability and good reliability, which must be linked to the inner statistical properties of the selected data: persistence, "fat tails" and good fit to mixtures of gaussian distributions.

[2] The latest start of the recession was dated by the NBER Committee on November 26, 2001, that is 7 months after the official peak. In the same way, the Committee dated the trough of the recession on July 17, 2003 that is one and a half year after the official date.

Over the past forty years, ten series, apart of the BCI ones, have sent robust signals to detect US recessions in real time and can lead to powerful and reliable recession models, whose structure should depend on the goal of the forecasters. Here, we particularly focus on three simple different models which should be associated to a univariate analysis. As described in Anas and Ferrara (2002) and Ferrara (2003), we rely a lot on the Industrial production, the Help Wanted Advertising Index, the Construction Spending and the Unemployment rate variables. We also identify other attractive "non-revised" data, which may definitely help to solve the puzzling "detection in real time" exercise. Those time series are deeply related to employment or consumer confidence: "Jobs Hard To Get, Present situation, Jobs judged Plentiful, Business condition judged Bad / Good". It is worth insisting that, in addition to be subject to recurrent revisions, three of four series of the BCI are among the fewer informative series, because of noisy disturbances or lagged signals.

More sophisticated models such as the DFMS or Stock and Watson's Experimental Recession Index seem not to be more powerful than simple (univariate or pseudo-multivariate) Hidden Markov Models. Thus the quality of information associated to the monitored time series seems to dominate the way of combining this information (and the complexity of related models).

Averaging data in temporal space seems to work marginally better than in probability space for monthly data, maybe thanks to favorable "pooling effects", that is a kind of neutralization of idiosyncratic shocks.

The paper is organized as follows: the next section discusses identification of reliable time series to track the growth and business cycles. Following suggestions of Chauvet and Pigger (2003) and the recent results of Anas an Ferrara (2002), the paper proposes in section 3 a careful study of the properties of those series with a Hidden Markov Model based on Hamilton (1989) as a monitoring tool of the American business cycle using a monthly data set. We also aim at challenging models with complex dynamic factor structure, as suggested by Ladiray (2002) and Watson (2001), by discussing the optimal and the simplest way of combining series through a probabilistic indicator. Section 4 offers comparison between famous other models like Stock and Watson (1991), the Dynamic Factor Markov with Switching Models of Chauvet (1998) or Kim and Nelson (1998) and a more basic but robust one like the "ISM rule of thumb"[3]. Section 5 discusses the particular question of "Real Time detection" and Section 6 concludes.

# 2. Some stylized facts about Business and Growth Cycles

## 2.1 The Data Set

The monthly data set covers the following period: February 1967 to March 2003. The sample was constructed in May 2003 and does not include any revision[4]. It incorporates 102 series among which the Conference Board Consumer Confidence Survey, the ISM (ex NAPM) survey, leading, coincident and lagging indicators of the Conference Board, monetary and financial data (spreads, monetary aggregates), diverse quantitative data (sector-based production indices, orders and shipments). We reconstructed some series of the semi-monthly Conference board Consumer Confidence Survey into monthly series by interpolation during the period 1967 – 1977.

[3] The US economy is said to enter a recession as the ISM Manufacturing Index breaks the 45 line.
[4] It is available on request such as the Gauss Programs.

On the whole part of this article, we systematically studied monthly growth rates through two different smoothing transformations:

$\mathrm{T}_k(L) = 1 - L^k$, with L the "Lag operator" applied to series in logarithms: $L^k X_t = X_{t-k}$.

Choosing k=1, that is using monthly growth rates as a high frequency transformation offers the earliest detection of a turning point but may present serious drawbacks of noisy disturbances. This is the reason why we used another smoothing degree k=3 which corresponds to a three month growth rate designed as a lower frequency transformation.

## 2.2 First insights throughout the literature and exploratory analysis

We first studied data properties through a simple Exploratory Data Analysis. Starting from a set of 102 series, we computed four indicators: spectral coherence and mean delay[5], time varying cross-correlation and applied an adaptation of the non-parametric algorithm "BBQ" from Bry and Boschan (1971), to assess linear and basic non linear relations between series and two monthly benchmarks: the Industrial Production Index and the NBER Dates. Before discarding any variable of our data set, we practiced an extensive review of the applied literature on leading and coincident indicators. This first "datamining work" led to a first qualitative ranking of 102 series that enabled us to achieve an initial classification among their leading and informative properties. The mean delay statistics enabled us to establish a qualitative sort of the series along their predictive horizon ability. We present those most informative candidates as "Cycle trackers" in Appendix A.

Leading and coincident indicators of the Conference Board have been extensively studied over the past twenty years. Stock and Watson (1989), Watson (2001), Anas and Ferrara (2002) gave a well documented review of their properties that we confirm in our initial work. The most advanced leading series (on average between 4 and 8 months) are the ten year-Treasury Bond / three month-T-bill spread and the Housing starts[6]. The leading indicator belongs to this category but suffers from an endogeneity bias: 50% of its weight is associated to the spread and the M2 aggregate, which reflect monetary conditions and financial expectations. Then come Stock prices which seem to be useful predictors of the growth cycle with a shorter lead, particularly one quarter ahead, as Estrella and Mishkin (1998) or Chauvet (1999) suggested. There are some series that also carry short advanced information (between 1 to 3 months) among which the well known Manufacturing Purchasing Manager Index of the ISM and its main components (Output, New Orders and Employment), some components of the Conference Board Consumer Survey (expectations), and some diverse quantitative data (Manufacturers new orders, consumer goods and material…). And at least, we exhibit some reliable coincident series among which we focused on for constructing our stochastic models.

[5] Following Croux and al (2001), we define the cross spectrum or spectral coherence of two time series X an Y as the ratio between spectral densities:

$C_{XY}(\omega) = \dfrac{f_{XY}(\omega)}{\sqrt{f_X(\omega) f_Y(\omega)}}$ that we can write on polar coordinates: $C_{XY}(\omega) = |C_{XY}(\omega)| e^{i\Phi(\omega)}$.

The Mean delay, which measures the advance / delay between series X and Y, is built as the ratio between the phase and the frequency :

$$M = \Phi(\omega)/\omega .$$

A positive Mean delay, for a given frequency, signals that the cyclical component of Y leads the cyclical component of X.

[6] Although this indicator has marginally lost its predictable power over the last two decades.

# 3. Monitoring the cycle through a simple HMM

The main innovations on that part of this paper is our peculiar focusing on time series that are non subject to revisions which were extracted from the Surveys of the Conference Board and the ISM. From a simple graphical analysis, strong properties of non-revised series appeared quite clearly such as the Help Wanted Advertising Index (HWA) as first noticed by Stock and Watson (1991) or Anas and Ferrara (2002) (see Figure 1). Some answers of the Conference Board surveys such as the "Jobs plentiful" (reversed) (see Figure 2) or "Business Conditions Judged Good" (see Figure 11 in Appendix A) answers also give no false alarm and sturdy coincident signals of a starting recession. Those series carry two kind of information: first the date of the turning point, second its intensity. For instance, the Figure 12 in Appendix A, associated to the unemployment rate, suggest two phases during the latest slowdown: a swift and violent contraction during the first months of the recession, a stabilization that meant the end of the recession and a new acceleration with a slower pace which can be related to the latest episode of the "1992" so called "jobless recovery" [7]. However, as asserted by Chauvet (1998), this classical monitoring does provide neither a formal mathematical framework to measure the business cycle, nor a probability model to interpret the aggregate information extracted from those variables.

All the series we initially selected seem to be highly persistent in "growth regime", that is featuring long periods of steady rise, disturbed by swift collapses, which are highly coincident with recessions. This gives us a hint of asymmetric shifts among expansion and contraction phases. That stylized fact would justify using non-linear concepts such as Hidden Markov Models. Contrary to the imposed assumption of linearity in Factor models such as Stock and Watson (1989) or simple OLS models, the HMM does not imply a built-in symmetry which forces expansions and contractions to have the same magnitude, duration and amplitude. This "asymmetry" for many US series is also suggested by a graphical study of the distribution of growth rates. They seem to be characterized by fat tails and might be modeled throughout mixtures of distributions[8]. Figure 3 presents the empirical distribution of the "Jobs hard to get" monthly growth rates. It validates the hypothesis of asymmetry and fat tails. HMM through the assumption of a mixture of gaussian should answer to most of these criteria. However are such intuitions confirmed by statistical estimates?

## 3.1 The univariate model

To confirm our first selection of the most reliable time series to detect the business cycle, we applied the same strategy as developed in Anas and Ferrara (2002) using a simple HMM to classify our data set. The univariate model is the more basic HMM, except that we transformed data by computing standardized growth rates for two main reasons. First, this is the best way to extract comparable results in term of standard deviation and statistically significance. On top of that, this enables us to deal with different objects (quantitative data, business surveys, rates…) and to aggregate them in a convenient way as they are expressed on the same scale (percentage point of standard deviation). Of course this specification assumes a certain stability of mean growth rates and its standard deviation in out of sample, which was not challenged by different periods of estimation.

---

[7] We should bear in mind that the BEA challenged the NBER dates of the 1990's recession and considered that this jobless recovery was indeed a recession phase. The last quarter of 2002 and the first quarter of 2003 strangely mimic this episode as a very mild "double dip" for a few series.

[8] See Hamilton (1994) for a comprehensive study of mixtures of distribution and their relation to Markov models.

**Figure 1: Help Wanted Advertising Index**

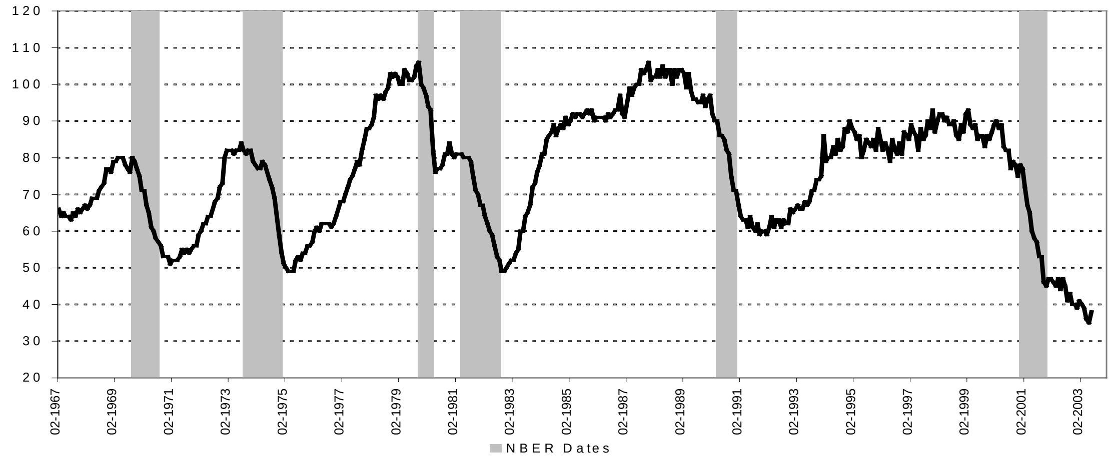



**Figure 2: Conference Board Consumer Survey: "Jobs Plentiful"**

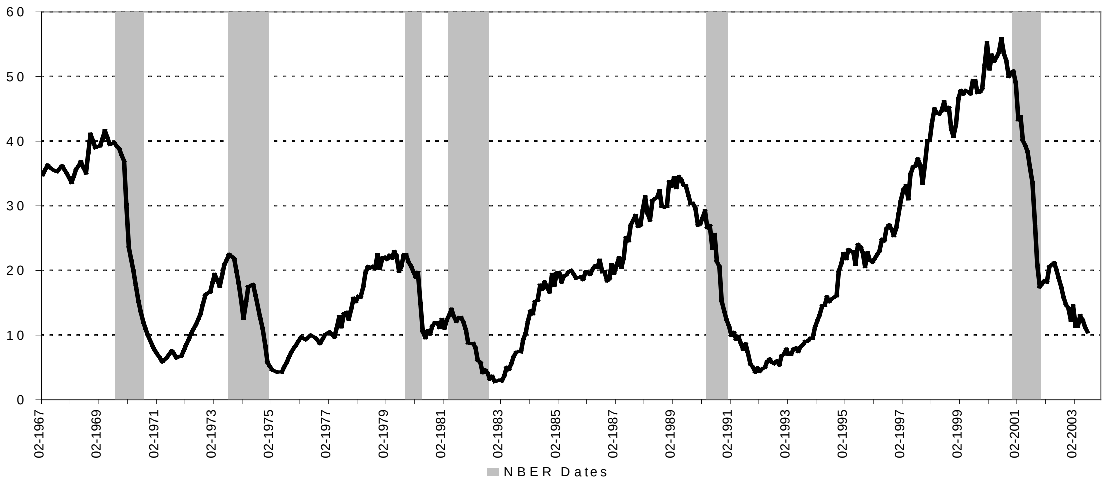
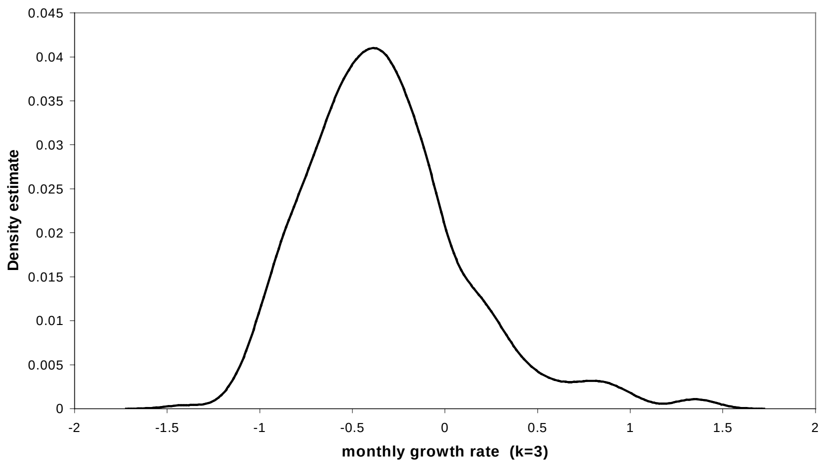



**Figure 3: Conference Board Consumer Survey: "Jobs Hard to Get" (reversed)**
**Kernel Density Plot – Estimation period: 1967-2 /1999-12**



We call $Y_i = (Y_{i,1}, \ldots, Y_{i,T})'$ the Tx1 vector associated to the time series i: and $y_i = (y_{i,1}, \ldots, y_{i,T})'$ the related vector of growth rates depending on k, that is the degree of smoothing ( k =1 or k= 3):

$$y_{i,t} = T_k(L)\log(Y_{i,t})\times 100 = \left(\log(Y_{i,t}) - \log(Y_{i,t-k})\right)\times 100$$

$y^*{}_i = (y^*{}_{i,1}, \ldots, y^*{}_{i,T})'$ corresponds to the vector of standardized growth rates:

$$y^*{}_{i,t} = \frac{(y_{i,t} - m_{y_i})}{\sigma_{y_i}}$$

with $m_{y_i}$ and $\sigma_{y_i}$ the corresponding empirical mean and standard deviation of growth rates of the series i. The model is then a simple modification of Hamilton (1989), which we remind the main features in Appendix B as a MS(2)-AR(0):

(1) $y^*{}_{i,t} = \mu_{S_t^i} + \varepsilon_{i,t} \qquad i = 1,\ldots,n$

The assumptions are:

(2) $\varepsilon_{i,t} \sim N(0, \sigma_i{}^2)$

(3) $\mu_{S_t^i} = \mu_0^i(1 - S_t^i) + \mu_1^i S_t^i$

where $\mu_0^i < \mu_1^i$ and $S_t^i = \{0,1\}$ is a “two states” unobserved variable, following a first order Markov Chain. When $S_t^i = 0$ ($S_t^i = 1$), the time series is said to be in a “contraction” (“expansion”) regime. The matrix of transition probabilities is defined very classically:

$$P\left(S_t^i = j / S_{t-1}^i = k,\ldots, S_1^i = l, I_{t-1}\right) = P\left(S_t^i = j / S_{t-1}^i = k\right) = p_{kj}^i$$

with k,j={1,0} and $I_{t-1} = \left(y^*_{i,t-1}, \ldots, y^*_{i,1}\right)$ defines the information set available in t-1.

In our basic specification, five parameters are estimated[9]:

$$\theta_i = \left(p_{00}^i, p_{11}^i, \mu_{0,}^i \mu_{1,}^i, \sigma_i^2\right)'$$

One may notice that the standard growth rates model can be written as a simple Hamilton’s model with a mild modification:

$$y_{i,t} = \tilde{\mu}_{S_t^i} + \eta_{i,t}$$

with $\tilde{\mu}_{S_t^i} = m_{y_i} + \sigma_{y_i}.\mu_{S_t^i}$ and $\eta_{i,t} = \sigma_{y_i}\varepsilon_{i,t}$ so that $\eta_{i,t} \sim N(0, (\sigma_{y_i}\sigma_i)^2)$.

Note that we also estimated this crude model:

$$y_{i,t} = \mu_{S_t^i} + \varepsilon_{i,t}\text{, where } \varepsilon_{i,t} \sim N(0, \sigma_i{}^2)$$

With no surprise we checked that the estimated transition probabilities and mean error indicators (see further for a definition of the relevant concepts) were perfectly identical. The only differences were linked to scale effects on mean and standard deviation estimates.

[9] And we have the well-known equality in relation to the Markovian transition probability matrix:

$$p_{00}^i + p_{10}^i = p_{01}^i + p_{11}^i = 1$$

A transformation of observations $\mathrm{T}_k(L) = 1 - L^k$ with a high value of k may reduce the variance and hence the false alarm probability. However the greater the degree k of smoothing, the bigger the risk of introducing spurious auto-correlation that may lessen the precision of the turning point dates and, as underlined in Andersson and al (2002), lower the probability of a successful detection. As suggested by Anas and Ferrara (2002) we applied their best smoothing degree (k=3) and followed Chauvet (1998) and Kim and Nelson (1998) by also applying (k=1) as the shortest delay of monthly detection, both for univariate and multivariate estimates.

We followed Albert and Chibb (1993), Lahiri and Wang (1994), Chauvet and Pigger (2003), Ferrara (2003) that assessed the performance of models with auto-correlation of order r ={0,1,2,3,4}.They find that the introduction of auto-correlation in the errors increases the risk of wrong inference concerning turning points. The effect of the auto-regressive parameters will largely be captured by the probabilities of remaining in the current state[10]. Besides, we favored one best practice of “parsimonious” and “flexible” modeling, by still bearing in mind the necessary tractability and sturdiness of our indicator.

## 3.2 Estimates and performances of the univariate monitoring

Our model with switching regimes is estimated by maximizing its likelihood function through classical optimization methods[11]. Many different samples of estimation were covered. The larger ones which encompass five to six recessions: 1967-2/1989-2, 1967-2/1999-12, 1967-2/2003-1 could have been too much influenced by the 70’s shocks. The shorter ones, 1979-1/1989-2, 1979-1/1999-12, 1979-1/2003-1, were designed to test the reliability of the series over a more recent phase but bear the drawback of a tiny set of months of recessions (between 30 to 38) which might deteriorate the statistical precision of the estimations. With no surprise, standard errors associated to these latest estimates are larger than those computed from the whole range period. Nevertheless, the estimated parameters depending on different samples remain quite close to each other. In addition to the classical computations on the “whole range”, we focused on periods on the brink of the two latest recessions: 1988-1993 and 1999-2003, to examine performance in the “neighborhood of the turning points”.

The usual criterion adopted to determine if the variable is in a recession state is whether the filtered probability of recession is greater than 50% :

$$P\left(S_t^i = 0 / I_t, \theta_i\right) > 0.5$$

Table 6 to 8 in Appendix B present the different estimation results of this univariate model applied to the thirteen best series we selected according to the Quadratic Probability Score (QPS) and the Absolute Probability Score (APS) criteria[12].

---

[10] Harding and Pagan (2001) provided a simple linear approximation of Hamilton’s model which let easily appear the already implicit “autocorrelation effect” associated with a Markov model.

[11] We applied also a Gibbs Sampling algorithm to check our results, especially for the DFMS model (see further), but this methods is still very costly in terms of computation time, especially for large data samples and ranges of estimates.

[12] The QPS for Quadratic Probability Score and APS for Absolute Probability Score related to the model i are defined as :

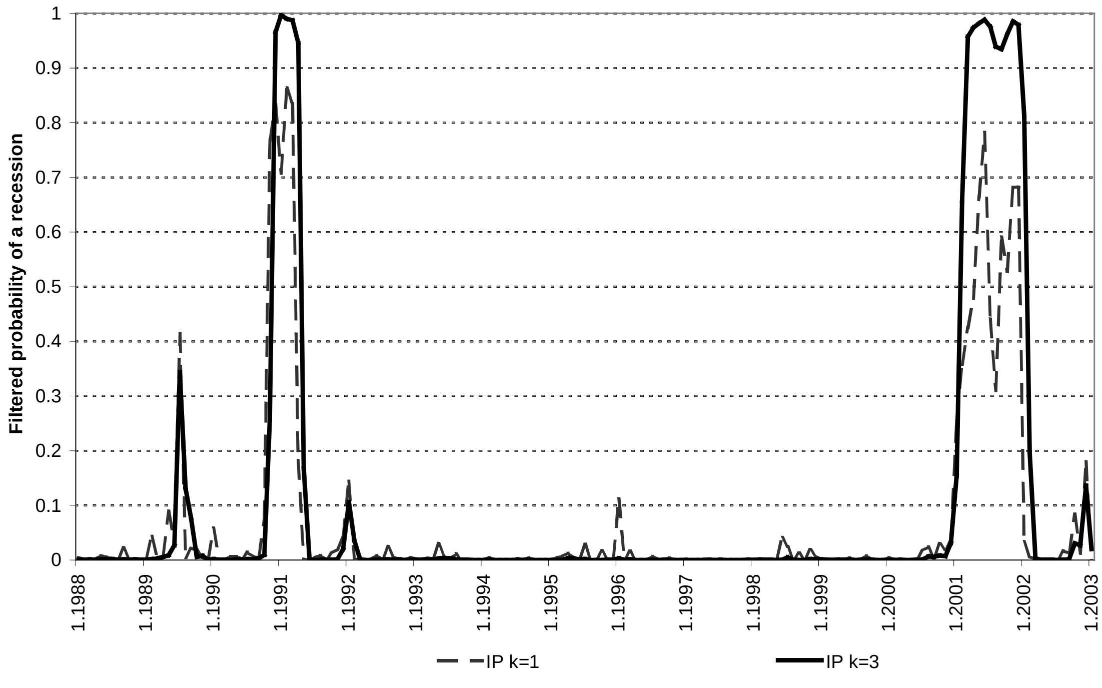
$$QPS(i,R) = \frac{1}{T}\sum_{t=1}^{T}\left(P_t^i - R_t\right)^2 \quad \text{and} \quad APS(i,R) = \frac{1}{T}\sum_{t=1}^{T}\left|P_t^i - R_t\right|$$

where $P_t^i$ designs the filtered Probability of being in recession and $R_t$ ={1,0} the dummy 1/0 dating the recession from the NBER committee.

This set of parameters was selected to judge the out-of sample performance of the model. As the QPS criterion, which is really close to the Mean Square Error concept, tends to more penalize bigger mistakes than the APS criterion does, we tend to favor this quadratic distance to establish our qualitative ranking. However we also computed other rules ("APS ranking", Geometric Average of QPS-APS) to check and challenge this classification. The results of the estimations seem to be quite stable and foolproof to different sampling period[13]. The expected duration of contraction of the best series appear quite close to the average recession duration which varies between 9 and 11 months depending on the reference period. This criterion could be determining to exclude series which are characterized by too short or too long phases of contraction[14]. On the contrary the estimated average expansion duration seem to be less reliable, around 55 months, whereas the last two cycles tended to be far longer (see Table 5 in appendix A to get the official NBER dates and related statistics).

We confirm that there may be large deterioration for some estimated parameters with high frequency data (k=1), whereas dealing with a higher smoothing degree such as k=3 makes a univariate monitoring quite an efficient tool, providing the series carry reliable and persistent information. There is however a trade-off: smoothing leads to a mild deterioration of detection delay (between one and two months) that may increase in real time. Figure 4 illustrates the deterioration of a HMM univariate estimate associated to the industrial production.

**Figure 4: Industrial Production HMM depending on different smoothing degrees (Out of sample 2000-1 / 2003-1)**




[13] Hamilton (1989) model, based on Quarterly GDP, failed to converge when applied to an extended range 1954-1995, whereas it was initially estimated on the 1954-1984 sample.
[14] Anas and Ferrara (2002) discarded the Personal Income, Manufacturing Sales and Employment series for these reasons.

Table 8 in Appendix A presents four different classification along the QPS criterion computed on the following estimating samples: 1967-2003, 1979-2003, 1988-1993 and 1999-2003. The ISM Purchasing Manager index “rule of thumb”[15], a non-revised series is presented as a benchmark for this univariate monitoring. We identified four series, confirming the results of Anas and Ferrara (2002), which seem to be particularly reliable and convey foolproof and global economic information through a simple HMM: the Help Wanted Advertising Index, the Unemployment rate, the Residential and Non residential Construction spending, the Industrial Production Index. Apart of these four indicators, the Business Coincident Index performs pretty well, but we should bear in mind that the performance of a “pure” real time detection should be lesser than this first pseudo-out of sample ranking. Some non-revised series such as “Business conditions judged Bad or Good”, “Hard to Get” or “Jobs Plentiful” have also carried reliable information over the past thirty years, but are on average a little less performing than our benchmark.

Monitoring the cycle with the univariate Hamilton’s model is with no doubt useful, but does not capture the notion of co-movements of sectoral variables. A first intuitive strategy would be to summarize the univariate filtered probabilities of a recession through a simple average, that is combining information in “Probability space”. But there are other alternatives to extract multivariate information.

# 4. Does a multivariate framework improve detection performances?

The empirical evidence in the forecasting literature, as reminded by Watson (2001) or Stock and Watson (2003), suggest that there may be huge pay-offs from “large” models approach, thwarting reported instability of forecasts based on individual leading indicators. However, the marginal predictive gain of including an additional regressor seems to be a sharply decreasing function of the number of predictors, a result that is shown to be consistent with a factor analytic structure linking regressions. Does this phenomenon hold for “Turning point” models? Following the insights of Diebold and Rudebusch (1996), Chauvet (1998) asserts the superiority of multivariate Markov switching models against the univariate framework. In this part we tend to challenge this position and discuss two main points: does an index of real activity constructed from a large number of variables perform better than a dynamical non linear factor model to detect turning points? Does averaging remain, as suggested by Watson (2001), an effective way of combining information to track recessions? If it does, how should data be aggregated?

## 4.1 “Pooled Data Markov Models” versus Dynamic Factor with Markov Switching Models

In this part we present a so-called “Pooled data” Model, that is a generalizing of the preceding univariate Hamilton’s model applied to an information summary (simple or weighted average, common factor on growth rates). We focus our attention on averages and factor structure to discuss the marginal gain associated to the implementation of more complex models such as the Dynamic Factor with Markov Switching (DFMS) model.

Starting form a (n x p) matrix of weights,

$$W = (W_1, \ldots W_p) ,$$

built according different criteria, enables us to test an extended array of models based on linear factor loadings, simple average, weighted average depending on inverted QPS or APS classifications, selected peculiar “loadings” for instance coming from Anas and Ferrara (2002)

[15] The Manufacturing ISM (ex NAPM) is said to detect a recession when it breaks the 45 line, see Figure 10 in Appendix A.

or a reconstructed Coincident index of the Conference Board. On top of that, the program is flexible enough to encompass as many series one would include. The “Pooled Data” Model is a simple Hidden Markov Model applied to a common factor built along “ad-hoc” information criteria. The framework of normalized growth rate lets us add together different signals (normalized monthly growth rates in our example)collected in the data space to build a new variable, as a representative of common information. We have the insights that, if series share the same quality of information, a simple average may perform as well as factorial models or dynamic factor models with Markov switching. There is some theoretical justification to use this kind of “aggregated information”. In fact, when several variables share the same change point, Wessman (1998) demonstrates that the minimal sufficient alarm statistic is univariate.

To build our (T x 1) vector of common information $x_j$, we extract a vector $W_j$ (n x 1) vector of fixed weights associated to the model j:

(4) $x_t^j = W_j^{\prime} y^*{}_t = \sum_{i=1}^{p} w_{i,j} y_{i,t}^* \text{ with } \sum_{i=1}^{p} w_{i,j} = 1.$[16]

(5) $x_t^j = \mu_{S_t^j} + \eta_t^j$

(6) $\eta_t^j \sim N(0, \omega_j^2)$

(7) $\mu_{S_t^j} = \mu_0^j (1 - S_t^j) + \mu_1^j S_t^j$

Table 1 presents the transposed W matrix and the associated “Pooled models”. Each lines corresponds to a weighing vector, expressed in percentage, related to a model and applied to our selected series.

**Table 1 :**

$W' = (W_1, \ldots, W_p)'$**, transposed Matrix of weights (expressed in percentage) for different “Pooled Data Models” or “Aggregated in Probability Space” Models.**

| | Coincident index | Unemployment rate | Help Wanted Adverstising Inde: | Industrial Production | Construction spendings | Condtions Bad | Jobs Hard To Get | Present situation | Jobs Plentiful | Conditions Good | Sales | Personal Income | Employed - non agricultural |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| QPS | 0.0 | 15.3 | 14.7 | 12.3 | 10.8 | 6.7 | 9.0 | 9.0 | 10.0 | 6.8 | 5.4 | 0.0 | 0.0 |
| APS | 0.0 | 12.1 | 10.7 | 11.7 | 10.6 | 8.5 | 9.3 | 10.8 | 10.7 | 8.2 | 7.5 | 0.0 | 0.0 |
| Mean | 0.0 | 10.0 | 10.0 | 10.0 | 10.0 | 10.0 | 10.0 | 10.0 | 10.0 | 10.0 | 10.0 | 0.0 | 0.0 |
| Mat1 | 0.0 | 100.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| Mat2 | 0.0 | 50.0 | 50.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| Mat3 | 0.0 | 33.3 | 33.3 | 33.3 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| Mat4 | 0.0 | 25.0 | 25.0 | 25.0 | 25.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| Mat5 | 0.0 | 20.0 | 20.0 | 20.0 | 20.0 | 0.0 | 0.0 | 0.0 | 20.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| Mat6 | 0.0 | 16.7 | 16.7 | 16.7 | 16.7 | 0.0 | 16.7 | 0.0 | 16.7 | 0.0 | 0.0 | 0.0 | 0.0 |
| Mat7 | 0.0 | 14.3 | 14.3 | 14.3 | 14.3 | 0.0 | 14.3 | 14.3 | 0.0 | 14.3 | 0.0 | 0.0 | 0.0 |
| Mat8 | 0.0 | 12.5 | 12.5 | 12.5 | 12.5 | 0.0 | 12.5 | 12.5 | 12.5 | 12.5 | 0.0 | 0.0 | 0.0 |
| Mat9 | 0.0 | 11.1 | 11.1 | 11.1 | 11.1 | 11.1 | 11.1 | 11.1 | 11.1 | 11.1 | 0.0 | 0.0 | 0.0 |
| Coinc | 0.0 | 0.0 | 0.0 | 25.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 25.0 | 25.0 | 25.0 |
| QPSMat4 | 0.0 | 28.8 | 27.7 | 23.2 | 20.3 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| QPSCoinc | 0.0 | 0.0 | 0.0 | 41.1 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 18.0 | 16.6 | 24.3 |
| QPSanas | 0.0 | 100.0 | 93.0 | 77.0 | 83.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| Nonrevis1 | 0.0 | 0.0 | 16.7 | 0.0 | 0.0 | 16.7 | 16.7 | 16.7 | 16.7 | 16.7 | 0.0 | 0.0 | 0.0 |
| Nonrevis2 | 0.0 | 0.0 | 25.0 | 0.0 | 0.0 | 0.0 | 25.0 | 25.0 | 25.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| NonrevisQPS | 0.0 | 0.0 | 26.2 | 0.0 | 0.0 | 11.9 | 16.1 | 16.0 | 17.7 | 12.2 | 0.0 | 0.0 | 0.0 |
| Factor | 8.6 | 7.8 | 8.1 | 8.2 | 6.1 | 7.7 | 7.7 | 8.4 | 8.4 | 7.8 | 7.0 | 6.5 | 7.7 |

This matrix W was also used as a key of aggregation to build weighted average of filtered probabilities (see further section 4.2). All models including the QPS or APS labels are built on normalized vectors so that :

$$w_{i,j} = \frac{\dfrac{1}{QPS(i,R)}}{\sum_{k=1}^{13} \dfrac{1}{QPS(k,R)} 1\{k \in W_j\}} \quad \text{or} \quad w_{i,j} = \frac{1}{\sum_{k=1}^{13} \dfrac{APS(i,R)}{APS(k,R)} 1\{k \in W_j\}},$$

[16] Except for the model QPSanas.

where the scores are extracted from prior univariate HMM estimates. The other models are built on simple averages (Mean, Mat1,…, Mat9, Coinc, Nonrevis1-2) or on loadings extracted from a principal component analysis (Factor).

We confront this simple model to more canonical specifications but which may be judged as over-parametrized ones such as the DFMS model estimated by Kim and Nelson (1998) as an adaptation of Chauvet (1998). A technical discussion is presented in Appendix C, dealing with relation between "Pooled Data HMM" and the DFMS model which the version we estimated is :

(8) $y^{*}_{i,t} = \lambda_i F_t + \upsilon_{i,t} \qquad i = 1,...,n$

(9) $\upsilon_{i,t} = \delta_i^1 \upsilon_{i,t-1} + \delta_i^2 \upsilon_{i,t-2} + \varepsilon_{i,t} \qquad i = 1,...,n$

(10) $F_t = \mu_{S_t} + \varphi_1 F_{t-1} + \varphi_2 F_{t-2} + \eta_t \qquad S_t = \{1,0\}$ following a first order Markov chain,

(11) $\mu_{S_t} = \mu_0 (1 - S_t) + \mu_1 S_t$

$\eta_t \sim N(0, \omega^2)$ and $\varepsilon_{i,t} \sim N(0, \xi_i^{\,2})$ mutually uncorrelated at all leads and lags.

Our modeling is only a short cut to build another "summary of information" with the main drawback of "adhocity" but clear advantage of simplicity, parsimony and tractability: whereas it costs only 5 parameters to be estimated whatever the number of exogenous variables may be, the DFMS would need from 25 parameters to estimate for the most parsimonious version, to 70 in an adapted version of Kim and Nelson (1998). We found it really difficult to obtain classical convergence for these models. For instance, we have failed to have the DFMS model converged on our five favorite series: IP, the HWA index, the Unemployment rate, "Jobs plentiful" and Construction spending whatever the estimation procedure may be[17]. However, this model is with no doubt attractive. It reconciles both the specifications of co-movements and the asymmetric shifts, extracts an unobserved common factor which may represent the business cycle and produces factor loadings which measure the sensitivity of each series to the business cycle. But we should bear in mind that the unobserved switching factor F remains a simple average of coincident and past growth rates, with dynamical weights. Because of its lack of parsimony and difficulty to obtain convergence, we tried to experiment simple but reliable challenging models. We also wanted to answer this main question: isn't it the inner quality of series which dominate rather the way of combining information?

The main results of our estimates are available in Tables 10 and 11 (see Appendix D). The presented selected pooled models[18] were built through a sequential analysis process. We first started from the ranking related to univariate HMM estimates, then add together one series at a time to build some new simple averages (Mat1,…, Mat9,…) or weighted averages according different criteria such as inverted QPS or APS[19]. For instance, Mat2 includes the unemployment rate and the HWA Index, Mat 4 adds IP and Construction Spending (the so-called "Anas and Ferrara series"), Mat5 includes the "Jobs plentiful component" in addition to these four series. We also computed two different "QPS averages" on Mat4 which appears to be our favorite model with Mat2 and Mat5.

[17] In addition to the Kim's Algorithm based on a Maximum Likelihood procedure and a linear approximation based on the Kalman Filter used by Chauvet (1998), we also estimated this specification with an adaptation of Gibbs sampling programs, but failed to obtain convincing estimates.

[18] We gave here some complementary details: a "Factor" is a normalized vector of factor loadings extracted from a principal component analysis which should be a good approximation of an optimal dynamic factor as proved in Doz and Lenglart (1999). We computed also "Pooled Data Models" built on non-revised series: Non revis1 and 2 are two different simple average, and NonrevisQPS is built on inverted QPS weights associated to the univariate estimates. "Coinc" is a simple average of the components of the Business Coincident Index monitored by the NBER Committee and challenging Chauvet's Filter.

[19] See prior Table 1.

Figure 5 gives evidence, as noticed by Chauvet (1998), that almost all recessions were preceded by an abnormal increase in filtered probabilities: "*Also a mild rise in probability 5 to 12 months before a recession forewarns a subsequent downturn, with the exception of the 1970 and 1975 recessions*". In fact most of the growth cycle turning points are signaled by our model: on February 1986, May 1989, April 1995 and May 2000 the filtered probability broke during more than two months the 5% borderline. These events are not considered as a recession due to their very short duration but send signals of an imminent slowdown that could turn into recession. All these brisk increases in filtered probability signal a stalled industrial activity and unfavorable employment conditions, which are quite coincident with an ISM index close to 50. The most noticeable spike occurred in May 2000, six months before the recession, which corresponded to a meaningful slowdown of GDP growth[20]. On the contrary, the models fail to detect the mild industrial slowdown which occurred in mid-1985. Over the past six recessions, a forecaster focusing on the "Pooled model 4" Mat4 would have been able to date a recession with mild delays (fewer than 2 months on average). Not to mention that, a weakened criterion of detection may lead to a better performance with not much risk of false signal. In fact every time the filtered probability broke the 30% line, a recession occurred in the two months to come.

Table 2 presents the official dates of the five latest recessions associated to each model and QPS results, using the strict 50% Hamilton's rule to date detections. To challenge the performance of the "Pooled Data Model", we present three benchmarks: the Stock and Watson Experimental Coincident Recession Index (XRI-C)[21], the model first presented in Chauvet (1998) and the basic ISM "rule of thumb.

**Table 2: Dating of US Business Cycle Turning points**
**NBER Reference, "ISM rule of thumb", Filtered probability of "Pooled Data" Models**

| | | Monthly advance*** / Leads (+) / Lags (-) over the NBER Dating procedure | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| *NBER Reference Dates* | | *Stock and Watson** | | *Chauvet (1998)** | | *ISM rule of thumb* | | *Mat 2** | | *Mat 4** | |
| Peak | Trough | Peak | Trough | Peak | Trough | Peak | Trough | Peak | Trough | Peak | Trough |
| December 1969 | November 1970 | -1 | 0 | -1 | 0 | -9 | 0 | 0 | -1 | 0 | 0 |
| November 1973 | March 1975 | -3 | -2 | -10 | -1 | -10 | -2 | -10 | -2 | -2 | -2 |
| January 1980 | July 1980 | -2 | 0 | -6 | 0 | -2 | 0 | -3 | 1 | -2 | 0 |
| July 1981 | November 1982 | -2 | 1 | -4 | 1 | -2 | 1 | -3 | 0 | -2 | 0 |
| July 1990 | March 1991 | -3 | 1 | -3 | 0 | -1 | 2 | -1 | 0 | -1 | 0 |
| March 2001 | November 2001 | 1 | 1 | -1 | 1 | 3 | 0 | 1 | 1 | 0 | 1 |
| Average | | -1.7 | 0.2 | -4.2 | 0.2 | -3.5 | 0.2 | -2.7 | -0.2 | -1.2 | -0.2 |
| Average 1980-2001 | | -1.5 | 0.8 | -3.5 | 0.5 | ***-0.5*** | ***0.8*** | ***-1.5*** | ***0.5*** | ***-1.3*** | ***0.3*** |

Sources: NBER July 2003 and Personal computations.
* Filtered probability.
** We don't take into account early warnings.
*** A signal is considered reliable if it persisted more than two months.

A simple strategy such as monitoring the Manufacturing Purchasing Manager Index seems to be quite efficient on average. However "Pooled models" have sent better qualitative and more persistent warnings. They also include series that cover a larger economic field (employment, IP, construction and manufacturing sectors) that can justify their primal using.

[20] Where as GDP growth was running at a 4% pace in annualized rate on the second quarter, it fell down to 0.0% during the third quarter of 2000 according the latest BEA estimates.
[21] Those series are available on Mark Watson's web site: http://www.wws.princeton.edu/~mwatson. We recall that contrary to Markov models, Stock and Watson probability model is not based upon a formal mathematical treatment, that could lead to breaking levels which signal an incoming recession. As a result the extracted filtered probability cannot be interpreted easily.

**Figure 5: "Pooled Data model 4", depending on IP, the Unemployment rate, the Help Wanted Advertising index and Construction spending**

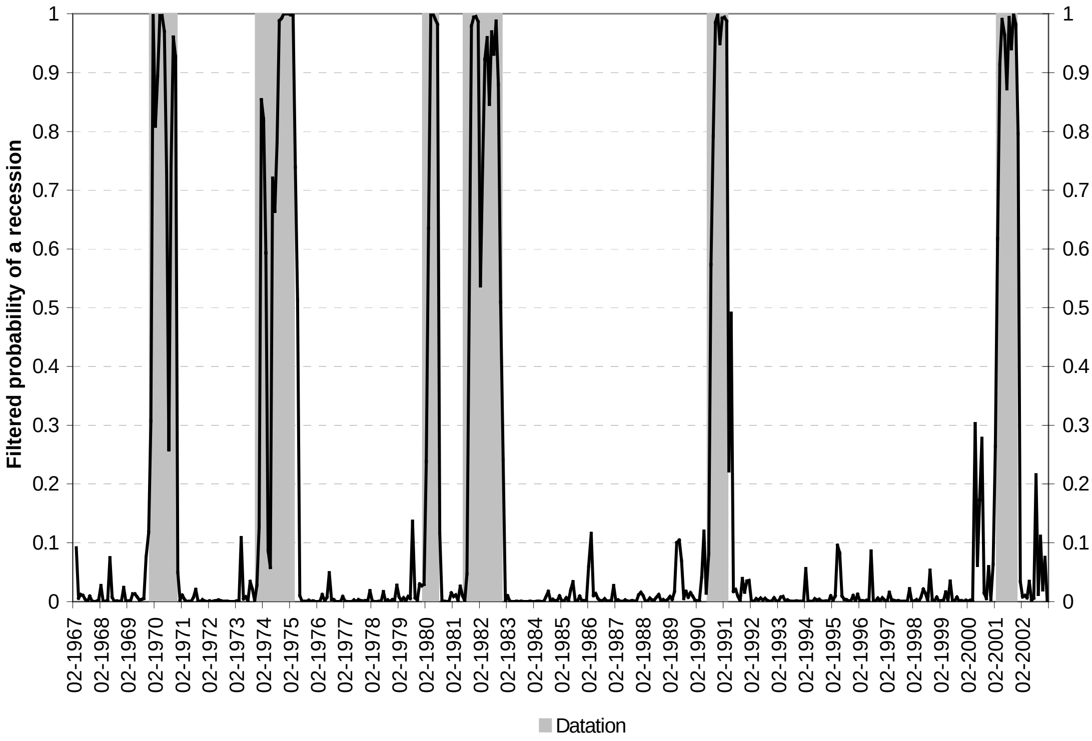
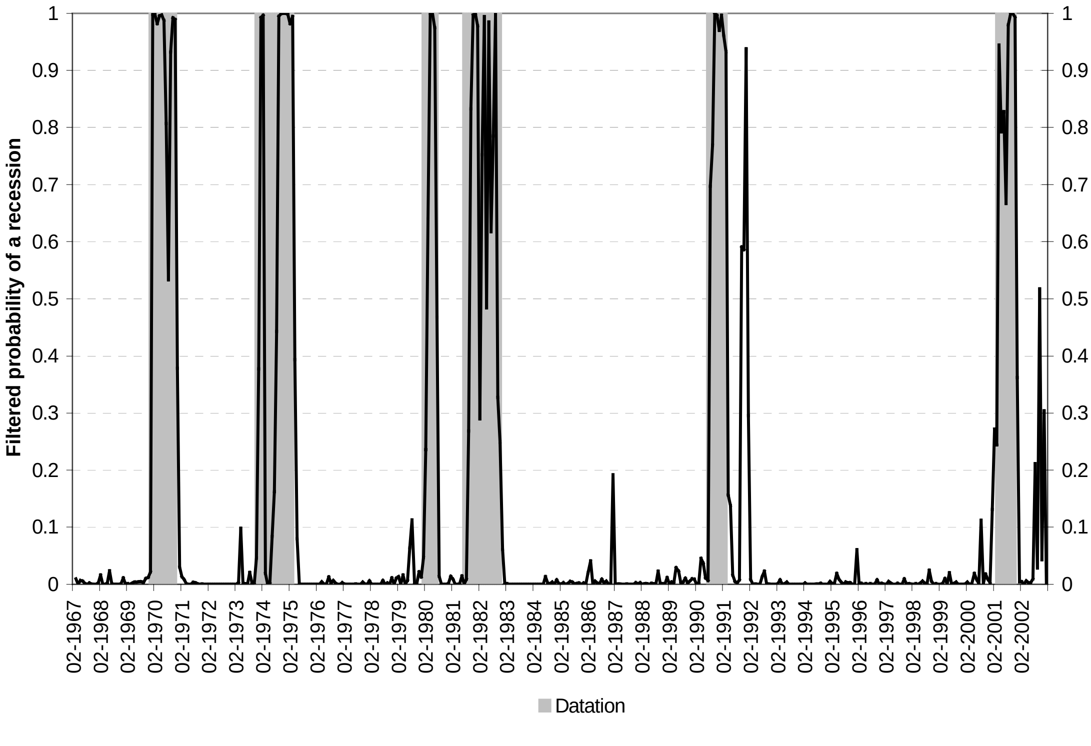



**Figure 6: "Pooled Data QPS" model :**
**Weighted average according to inverted univariate QPS criteria**



In addition to that, a systematic QPS ranking including many benchmarks presented in Table 12 in Appendix D tends to confirm that our best "Pooled Models" seem to be quite more reliable tools than the other benchmarks.

There does not seem to be any significant advantage for weighing the most reliable series[22], because the signal they convey might be robust enough. We also confirm the insights of Watson (2001): in our specification, there are not marginal gains to include more than four to six series (see Table 12 in Appendix D). However, tracking the cycle with an indicator based upon an exhaustive sample can send more reassuring signals to the forecaster. But there is a trade-off. Figure 6 shows that extending the number of variables in the sample may tend to delay the signal. Monitoring many series may then justify "factor modeling": as shown in Table 12 or in Figure 14 in Appendix D. In fact, simple averages including fewer reliable series (consider the "Mean" model) are outperformed by weighted average models such as the "QPS" model, that we favor as a more global indicator. Note also that factor loadings extracted from a principal component analysis are not more performing that simple averages.

## 4.2 Combining information in "Probability Space" or in "Temporal Space"

As emphasized by Carnot and Tissot (2002), the absence of theoretical arguments for combining data (in temporal space or in probability space) may be a caveat of Hidden Markov Models. Anas and Ferrara (2002) chose to combine information in probability space as a preferred solution, that is running HMM univariate estimations and aggregating the corresponding filtered probabilities through a single probability :

$$P\left(S_t^j = 0 / I_t, \theta_1 \ldots \theta_m\right) = \sum_{i=1}^{m} w_{i,j} P\left(S_t^i = 0 / I_t, \theta_i\right)$$

Yet, most of the "DFMS" literature favor the other approach: conducting a multivariate estimate (or a pseudo multivariate one as the "Pooled data models") leading to a single filtered probability computed through a prior combinations in temporal space. Using our weighing Matrix W as a key of aggregations, we ran the two type of models and discussed the performance according to different degrees of smoothing of both approach. If we rely on a simple statistical criterion inspired from the crude "Theil ratio" such as the relative "QPS ratio" and its related statistics reported in Table 3 we can draw some simple conclusions:

-"Pooled data" models tend to dominate averages of filtered probabilities that are extracted from univariate estimates. This must be linked to the poorer performance of some MS(2)-AR(0) estimates with too noisy signals. However, this result seems to be less robust for a higher smoothing degree (k=3).

-A qualitative analysis let us favor a combination of information through data spaces rather than a combination of filtered probabilities. Figure 7 illustrates that for a given transformation (smoothing degree k =3), a "Pooled Data Model" sent clearer and more stable signals than a model based upon the same weights and averaged filtered probabilities.

[22]Models based on QPS average such as QPSMat4 or built on the loadings proposed by Anas and Ferrara such as QPSanas lead to very marginal improvement in Quadratic Probability Scores.

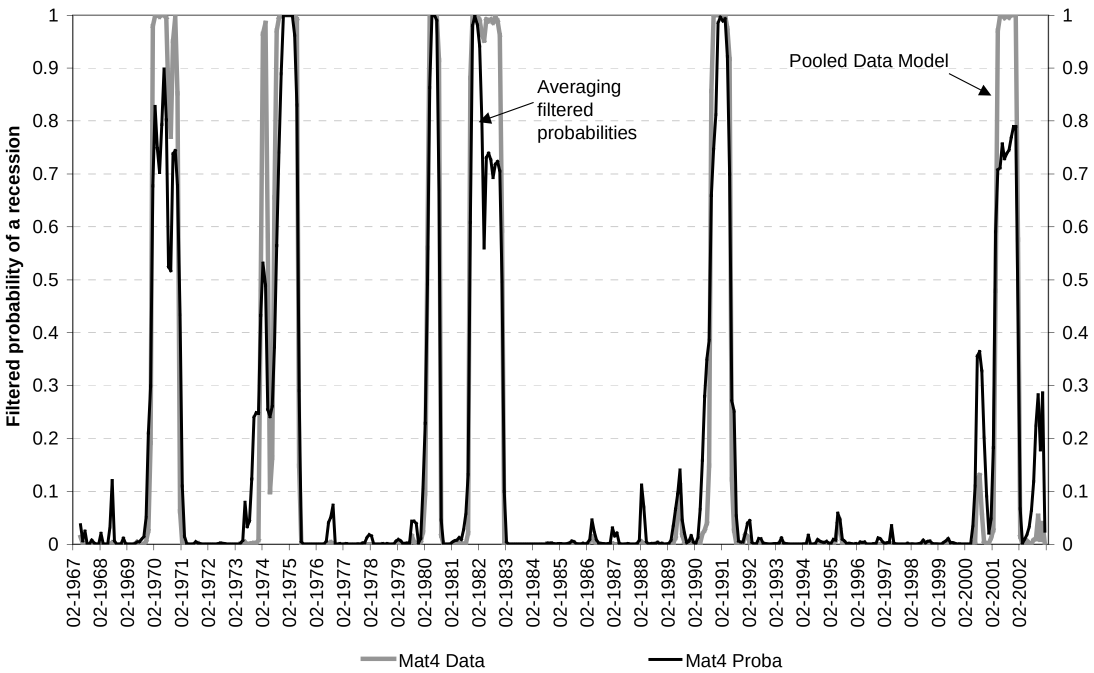
**Table 3:**
**Ratio: Pooled Data Models' QPS over Average of Filtered Probability Models' QPS**

| | 1967-2003 | 1979-2003 | 1988-1993 | 1999-2003* | 1967-2003 | 1979-2003 | 1988-1993 | 1999-2003* |
|---|---|---|---|---|---|---|---|---|
| | k=1 | k=1 | k=1 | k=1 | k=3 | k=3 | k=3 | k=3 |
| QPS | 0.74 | 0.71 | 1.00 | 0.42 | 1.05 | 0.97 | 1.50 | 0.55 |
| APS | 0.78 | 0.77 | 1.04 | 0.53 | 1.06 | 0.98 | 1.51 | 0.56 |
| Mean | 0.77 | 0.78 | 0.98 | 0.59 | 1.07 | 0.99 | 1.50 | 0.64 |
| Mat2 | 0.94 | 0.84 | 0.82 | 0.84 | 0.98 | 0.87 | 1.27 | 0.43 |
| Mat3 | 0.91 | 0.89 | 0.96 | 0.78 | 1.07 | 1.10 | 1.28 | 0.55 |
| Mat4 | 0.66 | 0.63 | 0.67 | 0.48 | 0.98 | 1.04 | 1.31 | 0.73 |
| Mat5 | 0.72 | 0.66 | 0.66 | 0.34 | 1.00 | 1.09 | 1.25 | 0.78 |
| Mat6 | 0.74 | 0.65 | 0.81 | 0.38 | 1.01 | 0.92 | 1.11 | 0.51 |
| Mat7 | 0.80 | 0.79 | 1.15 | 0.56 | 0.98 | 0.85 | 1.12 | 0.49 |
| Mat8 | 0.88 | 0.90 | 1.24 | 0.63 | 1.00 | 0.85 | 1.20 | 0.51 |
| Mat9 | 0.84 | 0.84 | 1.16 | 0.64 | 1.09 | 1.03 | 1.64 | 0.73 |
| Coinc | 0.74 | 0.90 | 1.21 | 0.58 | 0.86 | 1.08 | 1.11 | 0.98 |
| QPSMat4 | 0.69 | 0.65 | 0.66 | 0.53 | 1.03 | 1.05 | 1.33 | 0.68 |
| QPSCoinc | 0.83 | 0.90 | 0.99 | 0.59 | 0.87 | 1.04 | 1.01 | 0.73 |
| QPSanas | 0.07 | 0.06 | 0.04 | 0.05 | 0.07 | 0.06 | 0.06 | 0.04 |
| Nonrevis1 | 0.97 | 1.00 | 1.08 | 1.06 | 1.00 | 0.99 | 1.36 | 0.80 |
| Nonrevis2 | 1.03 | 1.00 | 1.39 | 1.03 | 1.03 | 1.00 | 1.36 | 0.75 |
| NonrevisQPS | 0.94 | 0.94 | 1.16 | 0.86 | 1.01 | 1.01 | 1.45 | 0.80 |
| Factor | 0.76 | 0.72 | 1.08 | 0.48 | 1.10 | 1.08 | 1.48 | 0.74 |
| Statistics : | | | | | | | | |
| Average | 0.78 | 0.77 | 0.95 | 0.60 | 0.96 | 0.95 | 1.26 | 0.63 |
| Median | 0.78 | 0.79 | 1.00 | 0.58 | 1.01 | 1.00 | 1.31 | 0.68 |
| Std deviation | 0.20 | 0.21 | 0.30 | 0.24 | 0.23 | 0.23 | 0.33 | 0.20 |
| First centile (1) | 0.69 | 0.64 | 0.66 | 0.37 | 0.87 | 0.85 | 1.09 | 0.48 |
| Last centile (2) | 0.95 | 0.95 | 1.21 | 0.89 | 1.07 | 1.08 | 1.50 | 0.80 |
| Inter centile ratio (1) /(2) | 0.73 | 0.68 | 0.54 | 0.42 | 0.81 | 0.79 | 0.73 | 0.60 |

*Pure out of sample estimates

*How to read this table: if the ratio is inferior to one, the Quadratic Probability Score of the Pooled Data Model (combination in Data Space) is better than the corresponding QPS of the averaging of filtered probabilities.*

**Figure 7: HMM aggregated in Probability Space or in Data Space**
**Smoothing transformation (k=3)**



# 5. Discussing the “Real Time detection” puzzle

Koenig, and al (2003) or Croushore and Stark (2003) argue that at each date within his sample, the econometrician estimating a forecasting equation ought to use only right-side data that would have been available at the time. The use of “real time data vintages “ was also deeply discussed to assess the performance of the leading indicator by Hamilton and Perez-Quiros (1996). This strategy may be objected whether the main goal of a research is to deem out (as in our article) the true and definitive qualitative relationships between the final release of time series and the dating of recession. Besides, real time data are seldom and difficult to gather: the data set of “Croushore and Stark” is unique and not comprehensive enough to cover our monthly sample.

However, we fully agree with the authors who criticized procedures using end of sample series which could be subject to large revisions. This is the reason why, as suggested by Stock (2003), we preferred a pseudo out of sample assessment, computing our estimates on a data set, reduced to the 1967-2 /1999-12 range which should be only marginally revised. Our strategy to respond to the “real time data vintage” puzzle was also to focus on the true performance of non-revised data.

As we formerly insisted, five series among ten in our selection are non-revised surveys, dealing mainly with the employment topic, which, according to Robert Hall[23], remains the most reliable source of information for the NBER Committee to date a recession. Those series carried reliable information in and out of sample and with real time estimates. The indicator presented in Anas and Ferrara (2002) includes two series subject to revisions among four (IP and Construction spending) but those series that we also include in our favorite “Pooled Data model” and simple index are far less revised than the other components of the Coincident indicator (Sales, and above all real Income or Employment) or quarterly data such as GDP. However, to avoid any caveat due to different vintages of IP and Construction spending, we computed pure non-revised indicators[24] from which our favorite model is the “Pooled model 2” made of the Unemployment rate and the HWA Index. We challenged them against the naïve and simple, yet not foolproof, rule of thumb of the Manufacturing “ISM” which usually calls the US economy to be in Recession every time the Purchasing Manager Index falls beyond 45.

As we mentioned it in the subsection 4.1, this “real time model” also sent early warnings of an impending downturn. As shown in Figure 8, the filtered probability reached the 10% level in may 2000, breaking even the 50% level during two months. This signal appeared one month before the peak of the S&P500 and the burst of the “Financial Bubble”. We can assume that those early warnings coming from the job market may have been carefully monitored by financial market operators who focus a lot on employment report data. The ISM rule signals the incoming industrial recession at the end of 2000, three months before the official start but with a lot of noise during the 8 months of the recession and a false alarm of recovery in August 2001.

Contrary to this latest episode, Figure 9 shows that the Purchasing Manager Index failed to give any early warning of recession in 1989. Our “non-revised” model exhibits small alarms during june 1989 and may 1990 and perform as well as the ISM, dating the recession in real time with a lag of one month.

[23] Robert Hall is the chairman of the NBER Dating Committee.
[24] See Non revis1, Nonrevis2 or NonrevisQPS loadings in section 4.1 which are other non revised indicators.

**Figure 8:**
**"ISM rule of thumb" and Pooled Model 2 during the 2001 recession**

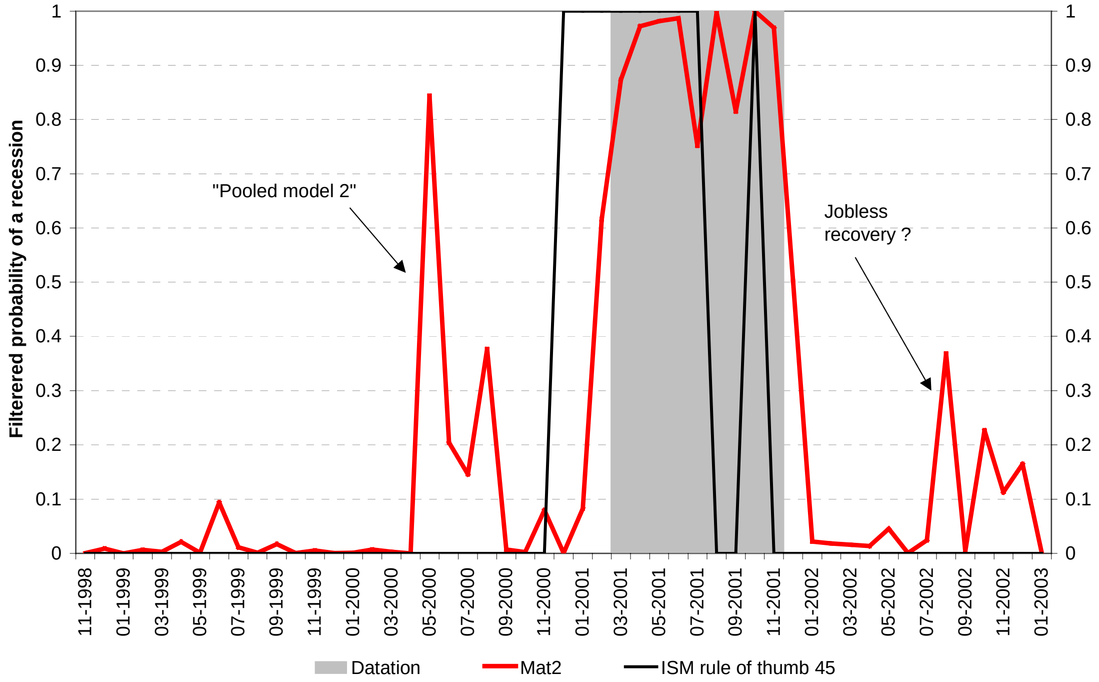
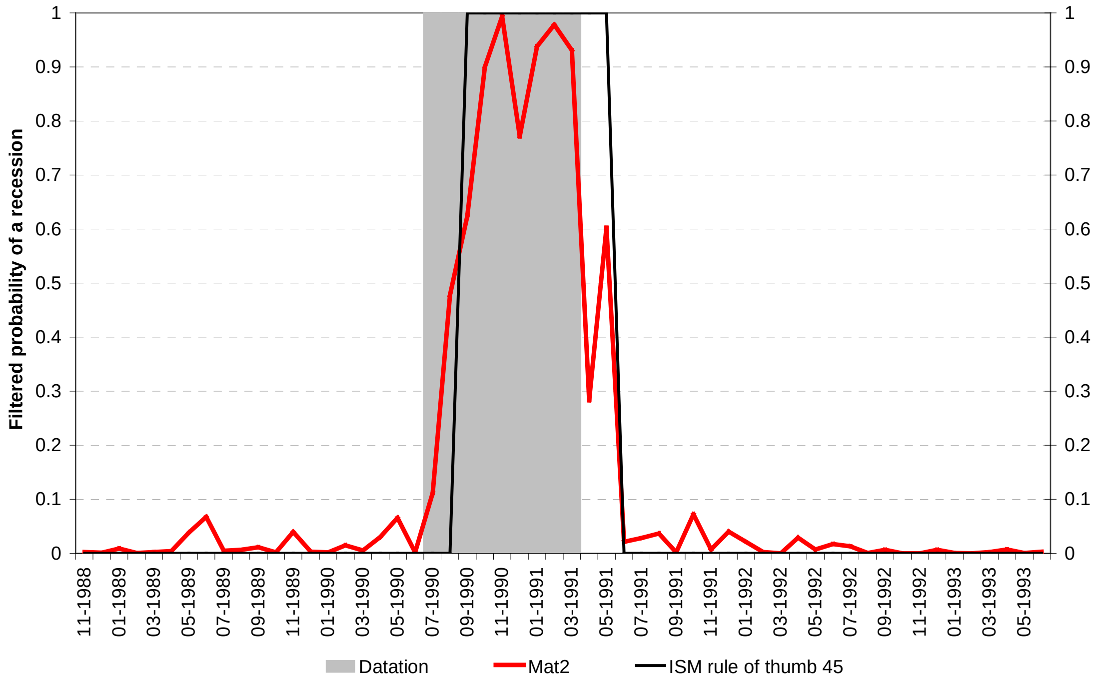



**Figure 9:**
**"ISM rule of thumb" and Pooled Model 2 during the 1990 recession**



## 6. Concluding remarks: what is worth it?

First and foremost, we exhibit ten series that carry a lot of information on the business cycle, four of which seem to be particularly reliable: the Unemployment rate, the Help Wanted Advertising Index, the Industrial Production and Construction spending. On top of that, some qualitative series remain quite attractive among which answers related to the job market situation. Basic methods keep challenging sophisticated tools: the Manufacturing ISM Purchasing Manager Index always gave sturdy signal. The 45 and 50 breaking lines have remained sturdy yet not foolproof rules of thumb to track business and growth cycles.

The models we exhibited are not "optimal" but represent a good trade-off between high reliability, wide coverage of the economy, tractability and robustness to different data vintages. We showed that simple strategies such as: monitoring those 13 series through a univariate HMM study, focusing on "Pooled data" Models[25] and still following simple rules such as the ISM Purchasing Manager Index, should help detecting growth slowdowns and early warnings of recession in real time. We also underlined that there are marginal gains, for the economic forecaster to build DFMS models whereas a simple adaptation of Hamilton's model perform as well, providing that he should choose the right way of transforming and combining data. However, there is no denying that models based upon a dynamic factor or a weighted average improve the reliability of signals especially when including series of poorer informative quality. These specifications also convey richer information by producing such a "common factor". However, the improvement in term of detection seems to remain marginal or null if a significant set of potent indicators is available.

Most of our work to build a reliable model was based on a craft sequential process. Adapting the Krolzig and Hendry (2001) algorithm to a non linear framework, by combining criteria of selection such as significance test and best QPS, APS or other distance criteria, might enhance the statistical performance of our selection. In the same way, we neglected the "multiple null hypothesis" puzzle when we sorted models through QPS criteria. In fact, we only tested a "one to one predictive ability" of each model. Extending to a non linear framework the "Data snooping" test presented in White (2000) or in Hansen (2001) could improve our results. However, Ladiray (2002) recalls that leading indicator modeling should not only be based upon statistical criteria but also reflect economic rationale. For instance, construction spending could have been excluded on pure statistical criteria during the latest recession, whereas it only reflected a modest and unusual slowdown in housing demand at this stage of the cycle.

We did not insist a lot in this paper on leading indicators. However, the same work was conducted to discuss leading properties of US time series, and was a little bit disappointing to identify lots of robust variables. Except the Leading Indicator Index, the Housing starts or the 10 year Treasury – 3 month T-Bill spread, few series seem to be able to give early and reliable signals over an horizon of six months. We also focused on German series (quantitative, surveys and financial leading indicators) with significant results which could lead to a further paper. Future work should focus on those series using new non parametric tools based on comprehensive study of distribution or HMM with a possible extended number of states such as Ferrara (2003). We did not explore Multivariate Qualitative Hidden Markov Models such as those developed by Gregoir and Lenglart (2000) or Baron and Baron (2002). However, this modeling which implements a supplementary stage of coding data may also be promising and might improve our first exploratory work.

[25] We favor the "Pooled Model 4" (Mat4), the "Pooled Model 2" (Mat2) for real time detection, and a weighted averaging model so-called "QPS" for global information (not to mention that the whole set of pooled models and univariate estimates remain complementary to have a look on).

# 7. References


**Albert, J H. and Chib, S**. (1993), "Bayes Inference via Gibbs Sampling of Autoregressive Time Series Subject to Markov Mean and Variance Shifts", *Journal of Business and Economic Statistics*, 11, pp. 1-15.

**Anas, J. and Ferrara, L**. (2002), "Un Indicateur d'Entrée et de Sortie de Récession: Application aux Etats-Unis", Centre d'Observation Economique, Paris, Working paper.

**Andersson, E. and Bock, D., Frisen, M**. (2002), "Detecting of Turning Points In Business Cycles", 26th CIRET Conference, Taipei , Working paper.

**Baron, H., and Baron, G**. (2002), "Un indicateur de retournement conjoncturel dans la zone euro", *Economie et Statistique*, 359-360, pp. 101-121.

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

**Burns, A F., and Mitchell, W E**. (1946), "Measuring Business Cycles." NBER, Columbia University Press.

**Carnot, N., and Tissot, B**. (2002), "La Prévision Economique", Economica (eds), pp. 120-128.

**Chauvet, M**. (1998), "An Econometric Characterization of Business Cycle Dynamics with Factor Structure and Regime Switching", *International Economic Review*, 39, pp. 969-96.

**Chauvet, M**.(1999), "Stock Market Fluctuations and the Business Cycle", Working paper.

**Chauvet, M**. (2002), "Is the U.S. Recession Over Yet?", *CREFC Economic Letter*, N°1.

**Chauvet, M. and Pigger, J M**. (2003), "Identifying Business Cycle Turning Points in Real Time", *Review of the Federal Reserve Bank of St. Louis*, pp. 47-62.

**Croushore, D., and Stark, T**. (2003), "A real Time Data Set for Macroeconomists: does the Data Set Vintage Matter?", *The Review of Economics and Statistics*, 85 (3), pp. 605-617.

**Croux, C., Forni, M., and Reichlin, L**. (2001),"A Measure of Comovement for Economic Indicators: Theory and Empirics", *The Review of Economics and Statistics*, 83, pp. 430-450.

**Diebold, F X., and Rudebusch, G D**. (1996), "Measuring Business Cycles : A Modern Perspective", *The Review of Economics and Statistics*, pp. 67-77.

**Doz, C., and Lenglart, F**. (1999), "Analyse factorielle dynamique: test du nombre de facteurs, estimation et application à l'enquête de conjoncture dans l'industrie ", *Annales d'Economie et Statistique*, N°54, pp. 90-125.

**Estrella, A., and Mishkin F S**. (1998), "Predicting U.S. Recessions: Financial Variables as Leading Indicators", *The Review of Economics and Statistics*, 80 (1), pp. 45-61.

**Ferrara, L**. (2003), "A Three-Regime Real-Time Indicator for the US Economy", *Economics Letters*, (81), Issue 3, December 2003, pp 373-378.

**Filardo, A J**. (1999), "How Reliable are Recession Prediction Models?", *Federal Reserve Bank of Kansas City*, *Quaterly Review*, Second quarter 1999, pp. 35-55.

**Frisen, M., and de Marè, J**. (1991), "Optimal Surveillance", *Biometrika*, 78, pp. 271-280.

**Gregoir, S., and Lenglart, F**. (2000), "Measuring the Probability of a Business Cycle Turning Point by Using a Multivariate Qualitative Hidden Markov model", *Journal of Forecasting*, n°19, pp. 81-102.

**Hamilton, J. D**. (1989), "A New Approach to the Economic Analysis of Non-stationary Time Series and the Business Cycle", *Econometrica*, 57 pp. 357-84.

**Hamilton, J. D**. (1994), "Time Series Analysis", Princeton University Press, Princeton, Chap 22.

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

**Hansen, P. R**. (2001), "An Unbiased and powerful test for superior predictive accuracy", *NBER Technical Working Paper*, 01-06, November 2001.

**Harding, D., and Pagan, A**. (2001), "A Comparison of Two Business Cycle Dating Methods", Working paper.

**Kim, M. J., and Yoo J. S**. (1995), "New Index of Coincident Indicators: a Multivariate Markov Switching Factor Model Approach", *Journal of Monetary Economics*,36, pp. 607-630.

**Kim, C. J., and Nelson, C. R**. (1998), "Business Cycle Turning Points, a New Coincident Index, and Tests of Duration Dependence Based on a dynamic factor model with regime switching.", *The Review of Economics and Statistics*, vol. 80, pp. 188-201

**Koenig, E. F., Dolmas S., Pigger J**. (2003), "The Use and Abuse of Real Time Data in Economic Forecasting." *The Review of Economics and Statistics*, 85 (3), pp. 618-628.

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

**Ladiray, D**. (2002), "Conjoncture, statistique et économétrie", *Economie et statistique*, 359-360, pp. 3-12.

**Lahiri, K., and Whang J. G**. (1994) "Predicting Cyclical Turning Points with leading index in the Markov Switching model.", *Journal of Forecasting*, vol. 13, pp. 245-263.

**Layton, A. P., and Smith D**. (2000), "Further on the three phases of the US Business Cycle", *Applied Economics*, vol. 32, pp. 1113-1143.

**Stock, J. H**. (2003), "How did Leading Indicator Forecast Do during the 2001 Recession?", NBER Working Paper.

**Stock, J. H., and Watson, M**. (1989), "New Indexes of Coincident and Leading Economic Indicators", *NBER Macroeconomics Annual*, The MIT Press, vol 4, pp. 351-394.

**Stock, J. H., and Watson, M**. (1991), "A Probability Model of the Coincident Economic Indicators", In *Leading Economic Indicators: New Approaches and Forecasting Records*, Cambridge University Press, pp. 63-89.

**Stock, J. H., and Watson, M**. (1993), "A Procedure for Predicting Recessions with Leading Indicators: Econometric Issues and Recent Experience", In *Business Cycles, Indicators and Forecasting*, J. Stock and M. Watson (eds), NBER, Chicago, pp. 255-284.

**Stock, J. H., and Watson, M**. (2003), "Forecasting Output and Inflation: The Role of Asset Prices", NBER Working Paper n° 8180, 2001, revised January 2003.

**Watson, M**. (2001), "Macroeconomic Forecasting Using Many Predictors", Working paper.

**Wessman, P**. (1998), "Some principle for Surveillance adopted for multivariate processes with a common change point", *Communication in Statistics*, Theory and Methods, 27, pp. 1143-1161

**White, H**. (2000), "A reality check for data snooping", *Econometrica*, (68) 5, pp.1097-1126.

# 8. Appendix

## Appendix A: Data Set and Selected Series

**Table 4: Most informative series depending on leading time horizon**

| LEADING HORIZONS | SERIES | SOURCE |
|---|---|---|
| *Four to eight months:* | | |
| | Interest rate spread, 10-year Treasury bonds less fed funds | DRI-Wefa |
| | Leading indicator | Conference Board |
| | New private housing units started | Bureau of Census |
| | USA Index of stock prices, 500 common stocks | Standard & Poors |
| *One to three months:* | | |
| | Money supply, M2 | Federal Reserve |
| | Consumer Confidence Survey : Expectations | Conference Board |
| | Average weekly hours | Bureau of Labor Statistics |
| | New claims for unemployment and insurances | Bureau of Labor Statistics |
| | ISM Purchasing managers index | ISM |
| | ISM Employment | ISM |
| | ISM New orders | ISM |
| | ISM Backlog of Orders | ISM |
| | Manufacturers new orders, consumer goods & mtrl | Bureau of Census |
| | Sales - manufacturing and trade, chainde 1996 dollars** | Bureau of Census |
| | Consumer Confidence Survey : Jobs hard to get* | Conference Board |
| | Consumer Confidence Survey : Jobs plentiful* | Conference Board |
| *Coincident :* | | |
| | Help Wanted Advertising Index** | Conference Board |
| | All manufacturing Industries - shipments | Bureau of Census |
| | Business Equipement Production | Federal Reserve |
| | Consumer Confidence Survey : Present situation* * | Conference Board |
| | Consumer Confidence Survey : Business conditions judged good** | Conference Board |
| | Consumer Confidence Survey : Business conditions judged bad* | Conference Board |
| | Coincident Index* | Conference Board |
| *Lagged by three to one months:* | | |
| | Manufacturing Prodution Index | Federal Reserve |
| | Unemployment rate** | Bureau of Labor Statistics |
| | Employed - Nonagricultural establishments (ESIT)* | Bureau of Labor Statistics |
| | Personal income less transfer payments chained 1996 dollars* | Bureau of Census |
| *Benchmark* | | |
| | Industrial Production Index** (Total Index) (G17) | Federal Reserve |

Data selected for the construction of stochastic indicators after a first univariate estimate of HMM are followed by *. Data belonging to the three final favored HMM are followed by **.

**Table 5: NBER Dating of US Recessions and Statistics**

| **BUSINESS CYCLE REFERENCE DATES** | | **DURATION IN MONTHS** | |
|---|---|---|---|
| **Peak** | **Trough** | **Contraction** | **Expansion** |
| *Quarterly dates are in parentheses* | | *Peak to Trough* | *Previous trough to this peak* |
| April 1960 (II) | February 1961 (I) | 10 | 24 |
| December 1969 (IV) | November 1970 (IV) | 11 | 106 |
| November 1973 (IV) | March 1975 (I) | 16 | 36 |
| January 1980 (I) | July 1980 (III) | 6 | 58 |
| July 1981 (III) | November 1982 (IV) | 16 | 12 |
| July 1990 (III) | March 1991 (I) | 8 | 92 |
| March 2001 (I) | November 2001 (IV) | 8 | 120 |
| Average, peacetime cycles: | | | |
| 1919-1945 (5 cycles) | | 10.0 | 52.0 |
| 1945-1991 (8 cycles) | | 10.3 | 49.3 |
| 1960-1991 (6 cycles) | | 11.2 | 54.7 |
| 1960-2003 (7 cycles) | | 10.7 | 64.0 |
| 1975-2003 (4 cycles) | | 9.5 | 70.5 |

Sources: NBER July 2003

**Figure 10: ISM Manufacturing Purchasing Manager Index "Rule of Thumb"**

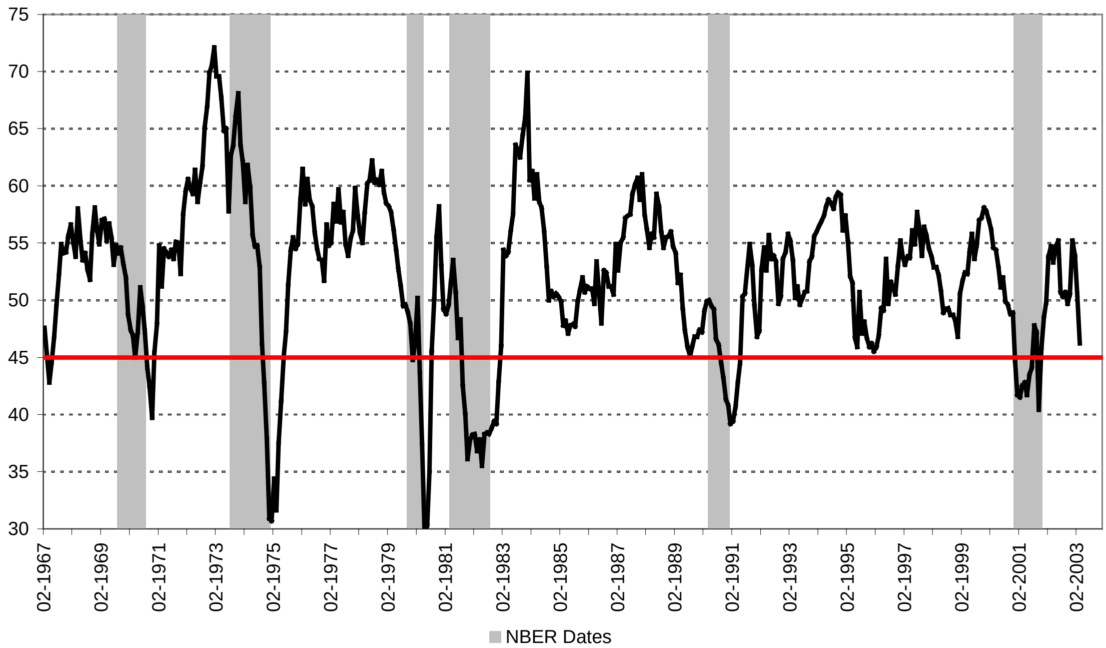


**Figure 11: Conference Board Consumer Survey: "Business Conditions Judged Good"**

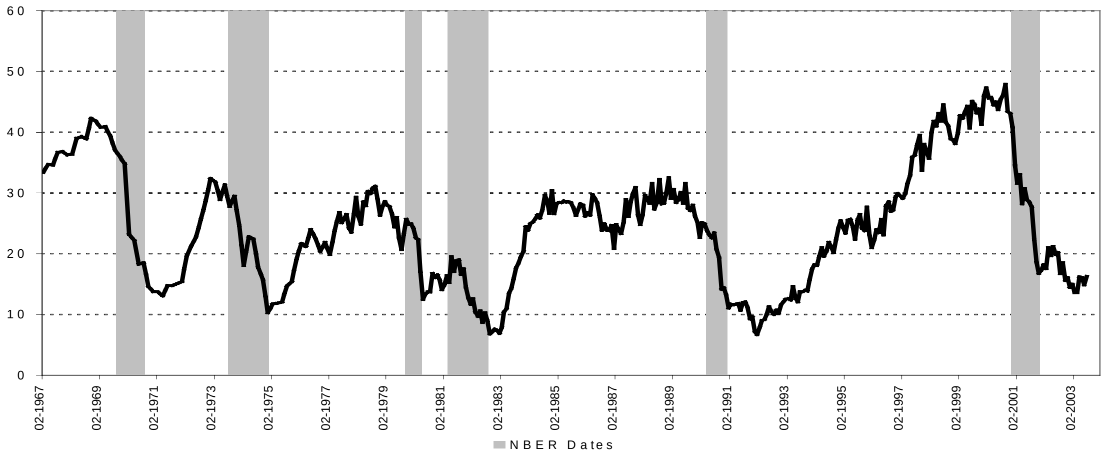



**Figure 12: Unemployment rate**

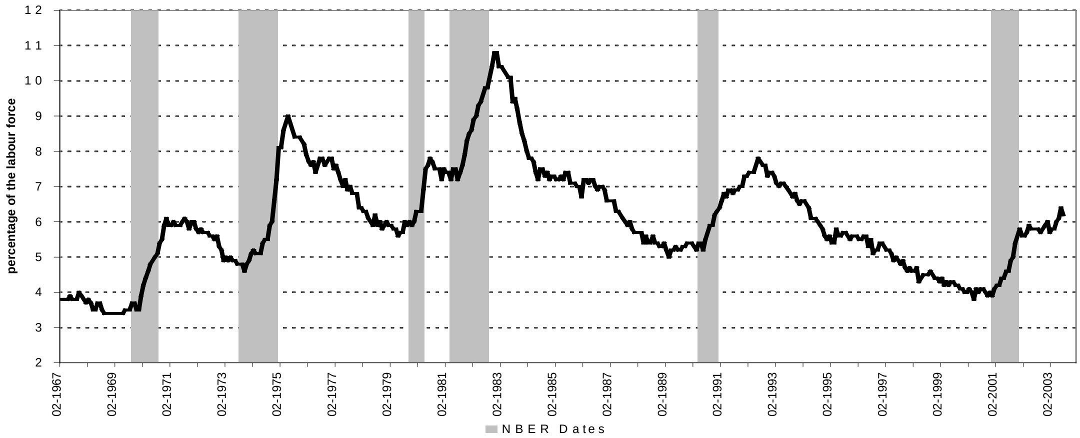
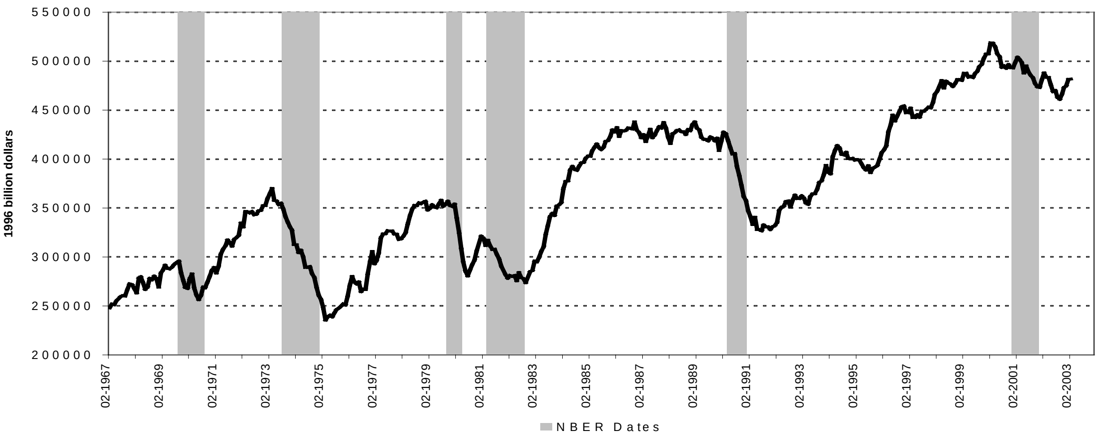



**Figure 13: Residential and Non-residential Construction Spending**



## Appendix B: The Univariate HMM, in-sample estimates and out of sample ranking

Hamilton's model is recursively estimated from the density built as a mixed-conditional distribution:

$$
\begin{aligned}
f\left(y_{i,t}^{*} / I_{t-1},\theta_i\right) &= \sum_{j=0}^{1} P\left(S_t^i = j, y_{i,t}^{*} / I_{t-1},\theta_i\right) \\
&= \sum_{j=0}^{1} P\left(S_t^i = j / I_{t-1},\theta_i\right) \times f\left(y_{i,t}^{*} / S_t^i = j, I_{t-1},\theta_i\right)
\end{aligned}
$$

It is in fact a mixture of gaussian distributions with the corresponding density:

$$
f\left(y_{i,t}^{*} / S_t^i = j,\theta_i\right) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{\left(y_{i,t}^{*} - \mu_j^i\right)^2}{2\sigma_i^2}\right), \quad j \in \{0,1\}
$$

The filtered probability of being for instance in a recession can then be seen, after some calculation, as a recursive function of a ratio of gaussian distributions, which is at the heart of the model based on the Theory of Optimal Surveillance:

$$
\begin{aligned}
P_t = P\left(S_t^i = 0 / I_t,\theta_i\right) = P\left(S_t^i = 0 / y_{i,t}^{*}, I_{t-1},\theta_i\right) &= \frac{P\left(S_t^i = 0, y_{i,t}^{*} / I_{t-1},\theta_i\right)}{f\left(y_{i,t}^{*} / I_{t-1},\theta_i\right)} \\
&= \frac{P\left(S_t^i = 0 / I_{t-1},\theta_i\right) \times f\left(y_{i,t}^{*} / S_t^i = 0, I_{t-1},\theta_i\right)}{\sum_{j=0}^{1} P\left(S_t^i = j / I_{t-1},\theta_i\right) \times f\left(y_{i,t}^{*} / S_t^i = j, I_{t-1},\theta_i\right)} \\
&= \frac{1}{1 + \left(\frac{1}{(p_{00} + p_{11} - 1) \times P_{t-1} + 1 - p_{11}} - 1\right) \times \frac{f\left(y_{i,t}^{*} / S_t^i = 1, I_{t-1},\theta_i\right)}{f\left(y_{i,t}^{*} / S_t^i = 0, I_{t-1},\theta_i\right)}}
\end{aligned}
$$

We also estimated the "Smoothed probability"[26], useful to date historically business cycles, but did not present its performance because of its poor detecting abilities in real time. The most noticeable properties of this model as shown in Hamilton (1989) is that the duration associated to the time spent by the variable in the state 1 conditional to a start in this regime follows a geometric distribution:

$$
P\left(D_t^i = k / S_1^i = 1\right) = \left(p_{11}^i\right)^{k-1}\left(1 - p_{11}^i\right)
$$

Consequently, we can calculate its first two moments, the expected duration (and its variance) of an expansion (recession) phases, really useful to build a diagnosis and estimate the reliability of a series:

$$
E\left(D_t^i / S_1^i = 1\right) = \frac{1}{1 - p_{11}^i} \text{ and } V\left(D_t^i / S_1^i = 1\right) = \frac{p_{11}^i}{\left(1 - p_{11}^i\right)^2}
$$

26 $P\left(S_t^i = j / I_T,\theta_i\right)$

We conducted estimates with two different smoothing transformations (k=1) and (k=3). We present here the estimation results for k=3, the best smoothing degree for univariate estimations. Asymptotic standard errors in parentheses correspond to the diagonal elements of the inverse Hessian obtained through numerical calculation. We recall that the expected duration of expansion (recession) are $1/(1-p_{11})$ and $1/(1-p_{00})$ .

**Table 6: The four most reliable series.**
**In-sample estimation results, 1967-2 / 1999-12 –480 observations**

| Parameters | Help Wanted advertising Index | Unemployment rate | Industrial Production Index | Construction Spending |
|---|---|---|---|---|
| $p_{11}$ | 0.9844 | 0.9819 | 0.984 | 0.9803 |
| | (0.007) | (0.0075) | (0.0071) | (0.0081) |
| $p_{00}$ | 0.9092 | 0.8918 | 0.8835 | 0.8885 |
| | (0.0377) | (0.0416) | (0.0496) | (0.0421) |
| $\mu_1$ | 0.7003 | 0.6752 | 0.7238 | 0.7345 |
| | (0.0255) | (0.025) | (0.0265) | (0.0271) |
| $\mu_0$ | -1.6749 | -1.7639 | -1.8246 | -1.5804 |
| | (0.1007) | (0.1001) | (0.1303) | (0.1129) |
| $\sigma^2$ | 0.3027 | 0.307 | 0.2596 | 0.2898 |
| | (0.0395) | (0.038) | (0.0405) | (0.0425) |
| Statistics: | | | | |
| QPS | 0.0491 | 0.0553 | 0.0782 | 0.0693 |
| APS | 0.0643 | 0.0688 | 0.0937 | 0.0905 |
| Average duration of: | | | | |
| Expansion | 64.1 | 55.2 | 62.5 | 50.8 |
| Recession | 11.0 | 9.2 | 8.6 | 9.0 |

**Table 7: The other most reliable and non-revised series**
**In-sample estimation results, 1967-2 / 1999-12 - 480 observations**

| Parameters | Business conditions judged Good | Business conditions judged Bad | Confidence index Present situation | Jobs hard to get | Jobs Plentiful |
|---|---|---|---|---|---|
| $p_{11}$ | 0.975 | 0.9796 | 0.9813 | 0.9791 | 0.9816 |
| | (0.0095) | (0.0084) | (0.0076) | (0.0085) | (0.0075) |
| $p_{00}$ | 0.8441 | 0.8297 | 0.8931 | 0.8763 | 0.9064 |
| | (0.0528) | (0.061) | (0.042) | (0.0486) | (0.0359) |
| $\mu_1$ | 0.7396 | 0.7551 | 0.6870 | 0.7136 | 0.6617 |
| | (0.0277) | (0.0281) | (0.025) | (0.0266) | (0.0241) |
| $\mu_0$ | -1.6527 | -1.8684 | -1.6982 | -1.6688 | -1.6460 |
| | (0.1279) | (0.1578) | (0.1141) | (0.1388) | (0.0947) |
| $\sigma^2$ | 0.2725 | 0.2287 | 0.3094 | 0.2926 | 0.3400 |
| | (0.043) | (0.0433) | (0.0394) | (0.0424) | (0.038) |
| Statistics: | | | | | |
| QPS | 0.0672 | 0.071 | 0.0781 | 0.0838 | 0.0896 |
| APS | 0.0879 | 0.091 | 0.0908 | 0.1046 | 0.1015 |
| Average duration of: | | | | | |
| Expansion | 40.0 | 49.0 | 53.5 | 47.8 | 54.3 |
| Recession | 6.4 | 5.9 | 9.4 | 8.1 | 10.7 |

**Table 8: Coincident index and its other components**
**In-sample estimation results, 1967-2 / 1999-12 - 480 observations**

| Parameters | USA Coincident Index | Sales manufacturing and trade | Personal income | Employed Non agricultural |
|---|---|---|---|---|
| $p_{11}$ | 0.9809 | 0.9694 | 0.9718 | 0.9823 |
| | (0,0095) | (0,0139) | (0,0131) | (0,0088) |
| $p_{00}$ | 0.9161 | 0.8852 | 0.8667 | 0.8986 |
| | (0,0389) | (0,0502) | (0,057) | (0,0473) |
| $\mu_1$ | 0.6462 | 0.7527 | 0.7039 | 0.6128 |
| | (0,0289) | (0,0363) | (0,0336) | (0,0273) |
| $\mu_0$ | -1.5345 | -1.2360 | -1.5031 | -1.8226 |
| | (0,1007) | (0,1454) | (0,2081) | (0,1373) |
| $\sigma^2$ | 0.377 | 0.3475 | 0.3331 | 0.3405 |
| | (0,0457) | (0,0608) | (0,0634) | (0,0449) |
| Statistics: | | | | |
| QPS | 0.0645 | 0.0813 | 0.1065 | 0.1231 |
| APS | 0.0799 | 0.1254 | 0.1361 | 0.1316 |
| Average duration of: | | | | |
| Expansion | 52.4 | 32.7 | 35.5 | 56.5 |
| Recession | 11.9 | 8.7 | 7.5 | 9.9 |

**Table 9: Ranking of univariate HMM**
**Smoothing degree k=3-In-sample parameter estimates 1967-2 / 1999-12**

| | 1967-2003 | | 1979-2003 | | 1988-1993 | | 1999-2003* | |
|---|---|---|---|---|---|---|---|---|
| | QPS | Rank | QPS | Rank | QPS | Rank | QPS | Rank |
| USA Coincident index | 0.0595 | 2 | 0.0567 | 4 | 0.0636 | 7 | 0.0674 | 4 |
| Unemployment rate | 0.0559 | 1 | 0.0432 | 1 | 0.0475 | 2 | 0.0426 | 1 |
| Help Wanted Adverstising Index | 0.0597 | 3 | 0.0454 | 2 | 0.0388 | 1 | 0.1365 | 8 |
| Industrial Production | 0.0782 | 5 | 0.0604 | 5 | 0.0778 | 8 | 0.0438 | 2 |
| Construction spending | 0.0848 | 8 | 0.0885 | 10 | 0.0592 | 6 | 0.1849 | 11 |
| Business Judged Bad | 0.0798 | 7 | 0.0765 | 8 | 0.0519 | 3 | 0.1239 | 7 |
| Jobs Hard To Get | 0.0956 | 12 | 0.0927 | 11 | 0.1081 | 11 | 0.1701 | 10 |
| Present situation | 0.0953 | 11 | 0.0944 | 12 | 0.1307 | 12 | 0.2148 | 14 |
| Jobs Plentiful | 0.1031 | 13 | 0.1009 | 13 | 0.1341 | 13 | 0.1958 | 12 |
| Business Judged Good | 0.0767 | 4 | 0.0790 | 9 | 0.0949 | 10 | 0.1394 | 9 |
| Sales Manufacturing and Trade | 0.0782 | 6 | 0.0698 | 6 | 0.0561 | 5 | 0.0837 | 5 |
| Personal Income | 0.0877 | 9 | 0.0738 | 7 | 0.0783 | 9 | 0.0507 | 3 |
| Employed - non agricultural | 0.1374 | 14 | 0.1303 | 14 | 0.1804 | 14 | 0.2742 | 15 |
| ISM inf 50 | 0.2168 | 15 | 0.2491 | 15 | 0.2778 | 15 | 0.2041 | 13 |
| ISM inf 45 | 0.0886 | 10 | 0.0554 | 3 | 0.0556 | 4 | 0.1224 | 6 |
| Statistics : | | | | | | | | |
| Average | 0.0932 | | 0.0878 | | 0.0970 | | 0.1370 | |
| Median | 0.0848 | | 0.0765 | | 0.0778 | | 0.1365 | |
| Std deviation | 0.0397 | | 0.0503 | | 0.0637 | | 0.0704 | |
| First centile (1) | 0.0596 | | 0.0494 | | 0.0493 | | 0.0466 | |
| Last centile (2) | 0.1237 | | 0.1185 | | 0.1619 | | 0.2105 | |
| Inter centile ratio (1) /(2) | 0.4815 | | 0.41663 | | 0.3043 | | 0.2212 | |

*Pure out of sample estimates

## Appendix C: The "Pooled Hidden Markov Model" (PHMM) and the Dynamic Factor with Markov Switching Model (DFMS)

The most general version of the DFMS may be written as :

$y^*_{i,t} = \lambda_i F_t + \upsilon_{i,t} \qquad i = 1...n$

$\delta_i(L)\upsilon_{i,t} = \varepsilon_{i,t} \qquad i = 1...n$, with $\delta_i(L) = 1 - \delta_{i,1}L \ldots - \delta_{i,p}L^p$

$\Phi(L)F_t = \mu_{S_t} + \eta_t \qquad S_t = \{1,0\}$ and $\Phi(L) = 1 - \phi_1 L \cdots - \phi_q L^q$

$\mu_{S_t} = \mu_0(1 - S_t) + \mu_1 S_t$

$\eta_t \sim N(0, \omega^2)$ and $\varepsilon_{i,t} \sim N(0, \xi_i^{\,2})$ mutually uncorrelated.

For simplicity, we focus our attention on the most parsimonious version of this model, that is :

$y^*_{i,t} = \lambda_i F_t + \upsilon_{i,t} \qquad i = 1...n$

$\upsilon_{i,t} = \varepsilon_{i,t} \qquad i = 1...n$

$F_t = \mu_{S_t} + \eta_t$

then, we can build from the DFMS our "Pooled Data model":

$$x_t^j = W_j' y^*_t = \sum_{i=1}^n w_{i,j} y^*_{i,t} = \sum_{i=1}^n w_{i,j}\lambda_i F_t + \sum_{i=1}^n w_{i,j}\upsilon_{i,t}$$

that is,

$$x_t^j = \sum_{i=1}^n w_{i,j}\lambda_i \mu_{S_t} + \sum_{i=1}^n w_{i,j}(\varepsilon_{i,t} + \lambda_i \eta_t) = \mu_{S_t^j} + \zeta_{j,t},$$

$\zeta_{j,t} \sim N(0, \sigma_j^2)$ with $\mu_{S_t^j} = \gamma_j \mu_{S_t}$ and $\sigma_j^2 = \sum_{i=1}^n w_{i,j}^2 (\xi_i^{\,2} + \lambda_i^2 \omega^2)$.

If the "true latent model" is a DFMS one, the qualitative "Markov" process extracted from a "Pooled Data Model" will be equivalent to the one extracted from the more sophisticated approach of the DFMS. However, the extraction of an unobserved factor to summarize the associated quantitative signal to the filtered probability, such as the coincident index does, remains the main advantage of DFMS models.

The most general version of this specification can be obtained by inverting the lagged polynoms (we assume that all the roots lie outside the unit-circle) and would get to a little more intricate equivalence such as:

$$x_t^j = \sum_{i=1}^n w_{i,j}\lambda_i \Phi^{-1}(L)\mu_{S_t} + \sum_{i=1}^n w_{i,j}(\delta_i^{-1}(L)\varepsilon_{i,t} + \lambda_i \Phi^{-1}(L)\eta_t) = \mu_{S_t^j} + \zeta_{j,t}$$

For instance, for p=q=1, the equivalent "Pooled Data Model" would be a mixture of MS(2)-AR(∞) :

$$\begin{aligned} x_t^j &= \sum_{i=1}^n w_{i,j}\lambda_i \sum_{k=0}^{\infty} \varphi^k L^k \mu_{S_t} + \sum_{k=0}^{\infty}\sum_{i=1}^n (w_{i,j}\delta_i^k \varepsilon_{i,t-k} + \lambda_i \varphi^k \eta_{t-k}) \\ &= \gamma_j \sum_{k=0}^{\infty} \varphi^k \mu_{S_{t-k}} + \sum_{k=0}^{\infty} \varphi^k \zeta_{j,t-k}, \quad \zeta_{j,t-k} = \sum_{i=1}^n \left(\frac{w_{i,j}\delta_i^k}{\varphi^k}\varepsilon_{i,t-k} + \lambda_i \eta_{t-k}\right) \end{aligned}$$

and $\zeta_{j,t-k} \sim N(0, \sigma_{jk}^2)$.

## Appendix D: Estimates of the « Pooled Hidden Markov Model » and its performance

Asymptotic standard errors in parentheses correspond to the diagonal elements of the inverse Hessian obtained through numerical calculation.

**Table 10: ”Pooled Data Models”**
**In-sample estimation results, 1967-2 / 1999-12 - 480 observations (k=1)**

| Parameters | Mat2* | Mat4 | QPSMat4 | QPSanas | Mat5 |
|---|---|---|---|---|---|
| $p_{11}$ | 0.9835 | 0.9835 | 0.9833 | 0.9834 | 0.9812 |
| | (0,0074) | (0,0074) | (0,0074) | (0,0074) | (0,0077) |
| $p_{00}$ | 0.8913 | 0.9067 | 0.9035 | 0.906 | 0.8865 |
| | (0,0454) | (0,04) | (0,0415) | (0,0404) | (0,0437) |
| $\mu_1$ | 0.6521 | 0.5592 | 0.559 | 1.9688 | 0.5172 |
| | (0,0239) | (0,0206) | (0,0206) | (0,0725) | (0,0191) |
| $\mu_0$ | -1.1815 | -1.0481 | -1.0664 | -3.6988 | -1.1161 |
| | (0,108) | (0,0888) | (0,0933) | (0,3163) | (0,0859) |
| $\sigma^2$ | 0.186 | 0.1941 | 0.1927 | 0.6835 | 0.1919 |
| | (0,0367) | (0,0321) | (0,0323) | (0,1133) | (0,0296) |
| Statistics: | | | | | |
| QPS | 0.0412 | 0.0280 | 0.0292 | 0.0279 | 0.0329 |
| APS | 0.0679 | 0.0519 | 0.0522 | 0.0513 | 0.0535 |
| Average duration of: | | | | | |
| Expansion | 60.6 | 60.6 | 59.9 | 60.2 | 53.2 |
| Recession | 9.2 | 10.7 | 10.4 | 10.6 | 8.8 |

* Signals an indicator made of non-revised series

**Table 11: “Pooled Data Models” - “Averaging effect”**
**In-sample estimation results, 1967-2 / 1999-12 – 480 observations (k=1)**

| Parameters | QPS | Mean | Coinc | NonrevisQPS* | Factor |
|---|---|---|---|---|---|
| $p_{11}$ | 0.9789 | 0.9779 | 0.9778 | 0.9782 | 0.9784 |
| | (0,0081) | (0,0084) | (0,0104) | (0,0087) | (0,0082) |
| $p_{00}$ | 0.8689 | 0.8611 | 0.8736 | 0.8667 | 0.8651 |
| | (0,0468) | (0,0494) | (0,0559) | (0,0563) | (0,0482) |
| $\mu_1$ | 0.4934 | 0.506 | 0.6049 | 0.6174 | 0.4879 |
| | (0,0184) | (0,0189) | (0,0249) | (0,0235) | (0,0182) |
| $\mu_0$ | -1.117 | -1.105 | -1.0728 | -1.1624 | -1.1329 |
| | (0,0794) | (0,0825) | (0,1035) | (0,1385) | (0,0819) |
| $\sigma^2$ | 0.1858 | 0.1815 | 0.1946 | 0.1962 | 0.1871 |
| | (0,0279) | (0,0284) | (0,0359) | (0,0364) | (0,0275) |
| Statistics: | | | | | |
| QPS | 0.0411 | 0.0466 | 0.0517 | 0.0603 | 0.0441 |
| APS | 0.0613 | 0.0679 | 0.0857 | 0.0881 | 0.0626 |
| Average duration of: | | | | | |
| Expansion | 47.4 | 45.2 | 45.0 | 45.9 | 46.3 |
| Recession | 7.6 | 7.2 | 7.9 | 7.5 | 7.4 |

* Signals an indicator made of non-revised series

**Table 12: Ranking of "Pooled HMM"**
**From in-sample estimates 1967-2 / 1999-12 –480 observations**

| | 1967-2003 | | 1979-2003 | | 1988-1993 | | 1999-2003* | |
|---|---|---|---|---|---|---|---|---|
| | QPS | Rank | QPS | Rank | QPS | Rank | QPS | Rank |
| QPS | 0.0430 | 6 | 0.0374 | 8 | 0.0396 | 10 | 0.0286 | 6 |
| APS | 0.0479 | 11 | 0.0440 | 11 | 0.0449 | 14 | 0.0401 | 13 |
| Mean | 0.0500 | 12 | 0.0468 | 16 | 0.0452 | 15 | 0.0466 | 17 |
| Mat2 | 0.0455 | 8 | 0.0320 | 6 | 0.0242 | 5 | 0.0391 | 12 |
| Mat3 | 0.0463 | 10 | 0.0341 | 7 | 0.0329 | 8 | 0.0304 | 7 |
| Mat4 | 0.0305 | 1 | 0.0246 | 2 | 0.0196 | 3 | 0.0236 | 3 |
| Mat5 | 0.0345 | 4 | 0.0282 | 4 | 0.0211 | 4 | 0.0193 | 1 |
| Mat6 | 0.0376 | 5 | 0.0297 | 5 | 0.0267 | 6 | 0.0235 | 2 |
| Mat7 | 0.0459 | 9 | 0.0406 | 10 | 0.0447 | 13 | 0.0378 | 10 |
| Mat8 | 0.0508 | 13 | 0.0481 | 17 | 0.0509 | 18 | 0.0454 | 16 |
| Mat9 | 0.0518 | 14 | 0.0482 | 18 | 0.0512 | 20 | 0.0483 | 18 |
| Coinc | 0.0548 | 16 | 0.0506 | 20 | 0.0553 | 21 | 0.0432 | 15 |
| QPS2 | 0.0320 | 3 | 0.0248 | 3 | 0.0193 | 2 | 0.0249 | 4 |
| QPSCoinc | 0.0590 | 18 | 0.0463 | 15 | 0.0442 | 12 | 0.0368 | 9 |
| QPSanas | 0.0306 | 2 | 0.0242 | 1 | 0.0191 | 1 | 0.0258 | 5 |
| USA Coincident index | 0.0590 | 19 | 0.0516 | 21 | 0.0590 | 23 | 0.0381 | 11 |
| Unemployment rate | 0.0592 | 20 | 0.0500 | 19 | 0.0297 | 7 | 0.0812 | 23 |
| Help Wanted Adverstising Index | 0.0585 | 17 | 0.0443 | 12 | 0.0469 | 16 | 0.0516 | 19 |
| Industrial Production | 0.0807 | 26 | 0.0584 | 23 | 0.0593 | 24 | 0.0560 | 21 |
| Construction | 0.0785 | 25 | 0.0770 | 27 | 0.0350 | 9 | 0.1617 | 33 |
| Confitions judged Bad | 0.1289 | 33 | 0.1264 | 34 | 0.1173 | 34 | 0.1544 | 31 |
| Jobs Hard To Get | 0.0906 | 29 | 0.0911 | 30 | 0.0619 | 25 | 0.1129 | 26 |
| Present situation | 0.0946 | 30 | 0.0929 | 31 | 0.0982 | 30 | 0.1326 | 29 |
| Jobs Plentiful | 0.0860 | 27 | 0.0858 | 29 | 0.0878 | 29 | 0.1289 | 28 |
| Confitions judged Good | 0.1270 | 32 | 0.1119 | 32 | 0.1169 | 33 | 0.1363 | 30 |
| Sales Manufacturing and Trade | 0.1371 | 34 | 0.1223 | 33 | 0.0991 | 31 | 0.1571 | 32 |
| Personal Income | 0.1702 | 35 | 0.1522 | 35 | 0.1389 | 35 | 0.1837 | 34 |
| Employed - non agricultural | 0.0982 | 31 | 0.0842 | 28 | 0.1105 | 32 | 0.1837 | 35 |
| Nonrevis1 | 0.0733 | 24 | 0.0713 | 26 | 0.0684 | 27 | 0.0934 | 25 |
| Nonrevis2 | 0.0689 | 23 | 0.0629 | 25 | 0.0773 | 28 | 0.0836 | 24 |
| NonrevisQPS | 0.0651 | 22 | 0.0606 | 24 | 0.0663 | 26 | 0.0667 | 22 |
| Factor | 0.0450 | 7 | 0.0383 | 9 | 0.0469 | 17 | 0.0325 | 8 |
| Stock & Watson | 0.0525 | 15 | 0.0460 | 13 | 0.0509 | 19 | 0.0426 | 14 |
| Chauvet | 0.0614 | 21 | 0.0460 | 14 | 0.0408 | 11 | 0.0522 | 20 |
| ISM inf 50 | 0.2168 | 36 | 0.2491 | 36 | 0.2778 | 36 | 0.2041 | 36 |
| ISM inf 45 | 0.0886 | 28 | 0.0554 | 22 | 0.0556 | 22 | 0.1224 | 27 |
| Statistics : | | | | | | | | |
| Average | 0.07223 | | 0.06492 | | 0.06342 | | 0.07748 | |
| Median | 0.05898 | | 0.04913 | | 0.05089 | | 0.04993 | |
| Std deviation | 0.04068 | | 0.04423 | | 0.04779 | | 0.05521 | |
| First centile (1) | 0.03605 | | 0.02893 | | 0.02264 | | 0.02536 | |
| Last centile (2) | 0.12795 | | 0.11707 | | 0.11369 | | 0.1594 | |
| Inter centile ratio (1) /(2) | 0.28178 | | 0.24715 | | 0.19918 | | 0.15907 | |

*Pure out of sample estimates

*How to read this table: Quadratic Probability Scores are computed starting with filtered probabilities estimated through a pseudo-out of sample process. We used the estimated parameters form 1967-2 / 1999-12 to compute in-sample and out of sample QPS. Then we sort this QPS and calculate diverse statistics.*

**Table 13:Dating US Business Cycle Turning points,**
**NBER, “ISM rule of thumb”, and Filtered probabilities of global average “Pooled Data” Models**

| | | Monthly advance*** / Leads (+) / Lags (-) over the NBER Dating procedure | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| *NBER Reference Dates* | | *ISM rule of thumb* | | *Coinc** | | *QPS** | | *Mean** | |
| Peak | Trough | Peak | Trough | Peak | Trough | Peak | Trough | Peak | Trough |
| December 1969 | November 1970 | -9 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| November 1973 | March 1975 | -10 | -2 | -3 | -1 | -2 | 0 | -2 | 0 |
| January 1980 | July 1980 | -2 | 0 | -3 | 0 | -2 | 1 | -2 | 1 |
| July 1981 | November 1982 | -2 | 1 | -3 | -1 | -3 | 2 | -3 | 2 |
| July 1990 | March 1991 | -1 | 2 | -3 | 1 | -1 | 0 | -1 | 0 |
| March 2001 | November 2001 | 3 | 0 | -1 | 1 | -1 | 0 | -1 | 0 |
| Average : | | -3.5 | 0.2 | -2.2 | 0.0 | -1.5 | 0.5 | -1.5 | 0.5 |
| Average 1980-2001 : | | -0.5 | 0.8 | -2.5 | 0.3 | -1.8 | 0.8 | -1.8 | 0.8 |

Sources: NBER July 2003 and Personal computations.
* Filtered probability.
** We don't take into account early warnings.
*** A signal is considered reliable if it persisted more than two months.

**Figure 14: Global averaged “Pooled Data models” during the 2001 recession**

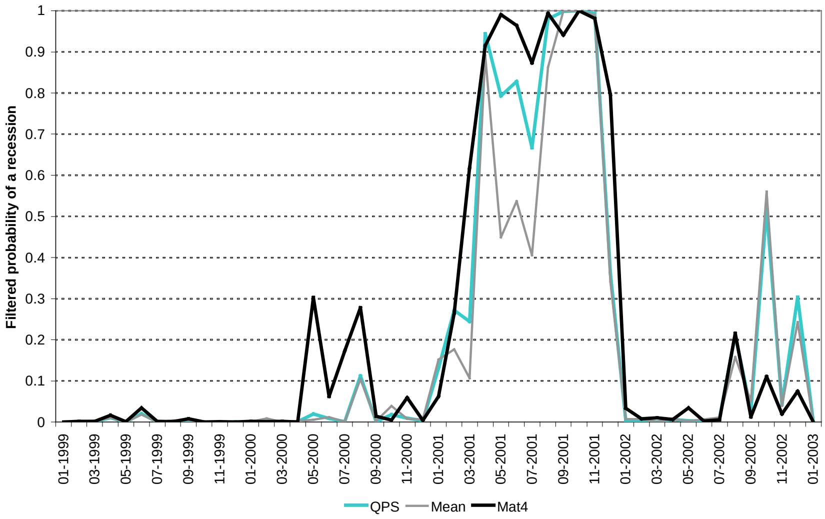



**Figure 15: Chauvet (1998) and “Pooled Data model” during the 2001 recession**

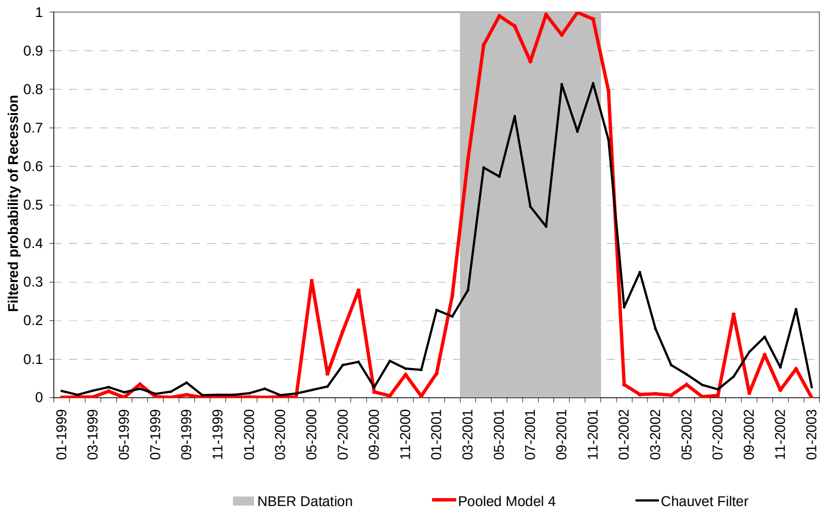


**Figure 16: Chauvet (1998) and "Pooled Data model 4"**

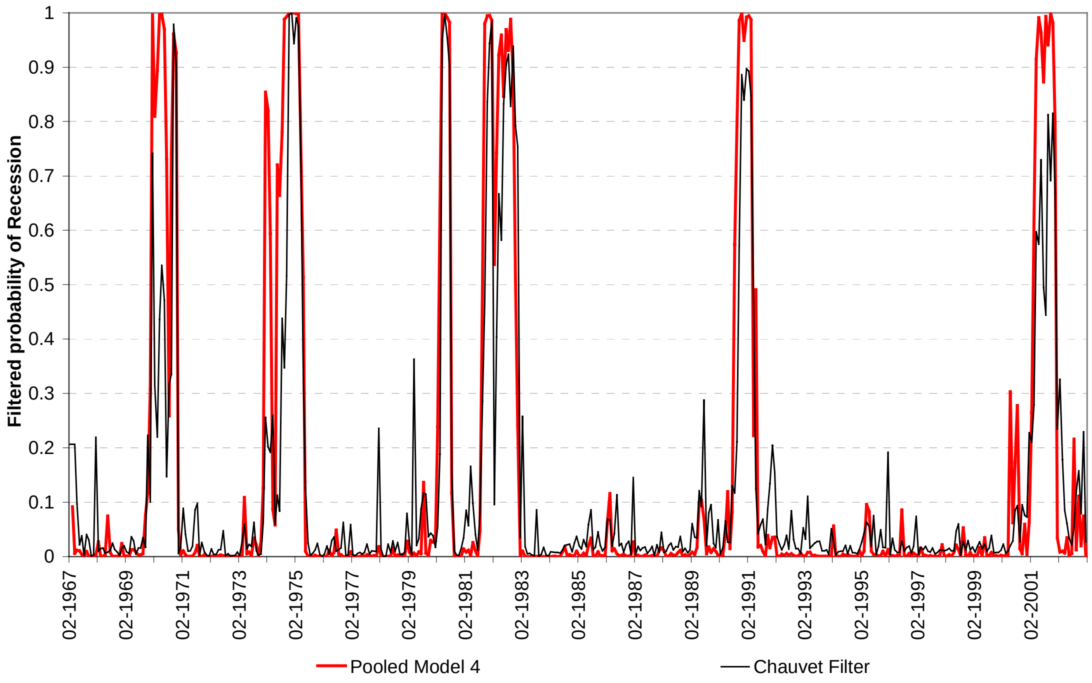



**Figure 17:**
**Stock and Watson ECI (1993) and "Pooled Data model 4" during the 2001 recession**