# Combination Schemes for Turning Point Predictions*


Monica Billio † Roberto Casarin†**
Francesco Ravazzolo‡¶ Herman K. van Dijk§‖

†University of Venice, GRETA Assoc. and School for Advanced Studies in Venice
‡Norges Bank
¶ BI Norwegian Business School
§Econometric Institute, Erasmus University Rotterdam
‖VU University Amsterdam and Tinbergen Institute



**Abstract**

We propose new forecast combination schemes for predicting turning points of business cycles. The combination schemes deal with the forecasting performance of a given set of models and possibly providing better turning point predictions. We consider turning point predictions generated by autoregressive (AR) and Markov-Switching AR models, which are commonly used for business cycle analysis. In order to account for parameter uncertainty we consider a Bayesian approach to both estimation and prediction and compare, in terms of statistical accuracy, the individual models and the combined turning point predictions for the United States and Euro area business cycles.



*JEL codes*: C11, C15, C53, E37.

*Keywords*: Turning Points, Markov-switching, Forecast Combination, Bayesian Model Averaging.



*We thank the referees and the editor for their very useful comments on an earlier version of our paper. We also thank the participants at: the 31st Annual International Symposium on Forecasting, 2011, Prague and the 5th CSDA International Conference on Computational and Financial Econometrics, 2011, London. The views expressed in this paper are our own and do not necessarily reflect those of Norges Bank.



**Corresponding author: Roberto Casarin, r.casarin@unive.it. Other contacts: billio@unive.it (Monica Billio); francesco.ravazzolo@norges-bank.no (Francesco Ravazzolo); hkvandijk@ese.eur.nl (Herman K. van Dijk).

# 1 Introduction

In recent years, interest has increased in the ability of business cycle models to forecast economic growth rates and turning points or structural breaks in economic activity. The early contributions in this stream of literature consider nonlinear models such as the Markov-switching (MS) models (see for example Goldfeld and Quandt [1973] and Hamilton [1989]) and the threshold autoregressive models (see Tong [1983] and Potter [1995]), both of which are able to capture the asymmetry and the turning points in business cycle dynamics. In this paper we focus on the class of MS models. We take the model of Hamilton [1989] as point of departure. For more recent data one needs an adequate business cycle model with more than two regimes (see also Clements and Krolzig [1998]) and a time-varying error variance. For example, Kim and Murray [2002] and Kim and Piger [2000] propose a three-regime (recession, high-growth, and normal-growth) MS model while Krolzig [2000] suggests the use of a model with regime-dependent volatility for the US GDP. In our paper we consider data on US and Euro industrial production, for a period of time including the 2009 recession and find that four regimes (high-recession, contraction, normal-growth, and high-growth) are necessary to capture some important features of the US and EU cycle in the strong-recession phases. As most of the forecast errors are due to shifts in the deterministic factors (see Krolzig [2000]), we consider a model with shifts in the intercept and in the volatility.

The first contribution of this paper is to exploit the time-varying forecast ability of linear and nonlinear models to produce potentially better forecasts. More specifically, in some empirical investigations and simulation studies, there is evidence that MS models are superior in in-sample fit, but not always in forecasting and that the relative forecast performance of the MS models depends on the regime present at the time the forecast is made (see for example Clements and Krolzig [1998]). It seems thus possible to obtain better forecasts by dynamically combining in a suitable way various model forecasts.

The second main contribution of this paper is to study the relationship between forecast combination and turning point extraction when many forecasts are available from different models for the same variable of interest. When many models are used for forecasting turning points, one can then alternatively combine the forecasts from the models and detect the turning points on the combined forecasts, or detect the turning points on the model forecasts and then combine the turning point indicators. We tackle this problem and show that the turning point forecasts are not invariant with respect to the order of the forecast combination and turning point extraction, and that the best combination should be evaluated in the specific case at hand. Our paper is related to Stock and Watson [2010], who consider the issue of dating the turning point for a reference cycle when many series are available. In this context, it is possible to detect clusters of turning points that are cycle-specific, and the problem of their aggregation becomes crucial to determine a reference cycle.

Another relevant contribution of the paper is a new model selection scheme which relies upon non-parametric measures, i.e. concordance statistics, of the proportion of time during which the predicted and the reference turning point series, are in the same state. The proposed scheme extends the literature on Bayesian model averaging (BMA) procedure (see Grunwald et al. [1993] for a review) for turning point forecasts. In the proposed approach to turning point forecast, we follow a Bayesian inference approach and account thus for both model and parameter uncertainty. The use of a Bayesian approach to forecast combination in business cycle analysis has been discussed in Min and Zellner [1993]. They consider both autoregressive (AR) models and AR models with time-varying parameters for predicting international output growth rates. Canova and Ciccarelli [2004] propose a Bayesian inference approach to the estimation of a multi-country panel model with time-varying parameters, lagged interdependencies and country specific effects. They follow Zellner et al. [1991] and predict turning points by using the predictive densities from their model. In this paper, we extend the previous literature and propose AR models with discontinuous (Markov-switching) dynamics in the parameters. The Bayesian approach proposed in this paper is based on a numerical approximation algorithm (Gibbs sampler) which is general enough to account not only for parameter uncertainty, but also for possible non-normality of the prediction error, as well as for nonlinearities of the data generating process. Another advantage of the Gibbs sampling procedures is that they naturally provide approximation of predictive density and forecast intervals for the variable of interest.

Finally, we study different strategies to specify combination weights. More specifically, we compare in terms of forecast performances weighting schemes driven by the prediction errors in predicting alternatively the level or the turning points of the variable of interest.

The paper is structured as follows. Section 2 introduces the Markov-switching model used in the analysis of the business cycle. Sections 3 and 4 present a Bayesian approach to inference and to forecast combination respectively. Section 5 provides a comparison between the forecasting methods for the Euro area and the US business cycles. Section 6 concludes the paper.

## 2 Predicting with Markov-switching Models

Let $y_t$, with $t = 1, \dots, T$, be a set of observations for a variable of interest. We consider two alternative autoregressive models for $y_t$. First, we assume that $y_t$ follows the Gaussian

AR process of order $p$, denoted with AR($p$),

$$y_t = \nu + \phi_1 y_{t-1} + \ldots + \phi_p y_{t-p} + u_t, \quad u_t \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2) \tag{1}$$

$t = 1, \ldots, T$, where $\nu$ is the intercept; $\phi_l$, $l = 1, \ldots, p$, are the autoregressive coefficients and $\sigma$ the volatility. In the following we will assume that the initial values, $(y_{-p+1}, \ldots, y_0)$, of the process are known. More generally, it is possible to include both the number of lag $p$ and the initial values in the inference process following, for example, the approach given in Vermaak et al. [2004] for the Gaussian AR processes.

Secondly, we consider a Gaussian AR process with parameters driven by a Markov-switching process and denote it with MS-AR. In an empirical study Clements and Krolzig [1998] present evidence that most part of forecast errors is due to time changes in some parameters of the prediction models. They suggested to consider, for example, MS models with regime-dependent volatility. In the present analysis, we follow Krolzig [2000] and Anas et al. [2008] and assume that both the intercept and the volatility are driven by a regime-switching variable. The resulting Markov-switching intercept and heteroschedasticity (MSIH) model, denoted with MSIH($m$)-AR($p$), is

$$y_t = \nu_{s_t} + \phi_1 y_{t-1} + \ldots + \phi_p y_{t-p} + u_t, \quad u_t \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2_{s_t}) \tag{2}$$

$t = 1, \ldots, T$, where $\nu_{s_t}$ is the MS-intercept; $\phi_l$, with $l = 1, \ldots, p$, are the autoregressive coefficients; $\sigma_{s_t}$ is the MS-volatility; and $\{s_t\}_t$ is the regime-switching process, that is a $m$-states ergodic and aperiodic Markov-chain process. This process is unobservable (latent) and $s_t$ represents the current phase, at time $t$, of the business cycle (e.g. contraction or expansion). The latent process takes integer values, say $s_t \in \{1, \ldots, m\}$, and has transition probabilities $\mathbb{P}(s_t = j | s_{t-1} = i) = p_{ij}$, with $i, j \in \{1, \ldots, m\}$. The transition matrix $P$ of the chain is

$$P = \begin{pmatrix} p_{11} & \cdots & p_{1m} \\ \vdots & & \vdots \\ p_{m1} & \cdots & p_{mm} \end{pmatrix}$$

and has, as a special case, the one-forever-shift model that is widely used in structural-break analysis (e.g., see Jochmann et al. [2010] and references therein). As for the AR case, in our applications we assume that the initial values, $(y_{-p+1}, \ldots, y_0)$, and $s_0$, of the processes $\{y_t\}_t$ and $\{s_t\}_t$ respectively, are known. A suitable modification of the procedure in Vermaak et al. [2004] can be applied for estimating the initial values of both the observable and the latent variables.

# 3 Bayesian Inference

## 3.1 Data Augmentation

In this paper we follow a Bayesian inference approach. One of the reasons of this choice, is that inference for latent variable models calls for simulation based methods, which can be naturally included in a Bayesian framework. Moreover, model selection and averaging can be easily performed in an elegant and efficient way within a Bayesian framework, overcoming difficulties of the frequentist approach in dealing with model selection for non-nested models.

In this paper we propose a Bayesian inference framework that relies on data augmentation (see Tanner and Wong [1987]) and on a Monte Carlo approximation of the posterior distributions. Following this approach, we introduce the allocation variable $\boldsymbol{\xi}_t = (\xi_{1t}, \ldots, \xi_{mt})$, where $\xi_{kt} = \mathbb{I}_{\{k\}}(s_t)$ indicates the regime to which the current observation $y_t$ belongs to, and $\mathbb{I}_A(x)$ is the indicator function that takes value 1 if $x \in A$ and 0 otherwise. The allocation variables cluster the observations in different groups. Each group corresponds to a regime and is characterized by regime-specific parameters in the regression equation. In the following, a configuration of the allocation variables such that at least one group has not a minimum number of observations is referred to as troublesome grouping. Secondly, we write the random-coefficient dynamic regression model in equation (2) as follows

$$y_t = \sum_{k=1}^{m} \xi_{kt}\nu_k + \phi_1 y_{t-1} + \ldots + \phi_p y_{t-p} + u_t, \quad u_t \overset{i.i.d.}{\sim} \mathcal{N}\left(0, \sum_{k=1}^{m} \xi_{kt}\sigma_k^2\right) \tag{3}$$

For reason of expository convenience, we follow Frühwirth-Schnatter [2006] and define the vector of regressors, $\mathbf{x}_{0t} = (y_{t-1}, \ldots, y_{t-p})'$, with regime invariant coefficients, $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_p)'$, and the two vectors, $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_m)'$ and $\boldsymbol{\sigma} = (\sigma_1^2, \ldots, \sigma_m^2)'$, of regime-specific parameters. In this notation the regression model in equation (3) writes as

$$y_t = \boldsymbol{\xi}_t'\boldsymbol{\nu} + \mathbf{x}_{0t}'\boldsymbol{\phi} + u_t, \quad u_t \overset{i.i.d.}{\sim} \mathcal{N}(0, \gamma_t)$$

where $\gamma_t = \boldsymbol{\xi}_t'\boldsymbol{\sigma}$ is the MS heteroschedasticity (or stochastic volatility) process.

The data-augmentation procedure, described above, yields the completed likelihood function

$$L(\mathbf{y}_{1:T}, \boldsymbol{\xi}_{1:T}|\boldsymbol{\theta}) = \prod_{t=1}^{T}\prod_{k=1}^{m}\prod_{j=1}^{m} p_{jk}^{\xi_{jt-1}\xi_{kt}} \left(2\pi\sigma_k^2\right)^{\frac{-\xi_{kt}}{2}} \exp\left\{-\frac{\xi_{kt}}{2\sigma_k^2}(y_t - \nu_k - \mathbf{x}_{0t}'\boldsymbol{\phi})^2\right\} \tag{4}$$

where $\boldsymbol{\theta} = (\boldsymbol{\nu}', \boldsymbol{\phi}', \boldsymbol{\sigma}', \mathbf{p})'$ is the parameter vector, with $\mathbf{p} = (\mathbf{p}_{1\cdot}, \ldots, \mathbf{p}_{m\cdot})'$, $\mathbf{p}_{k\cdot} = (p_{k1}, \ldots, p_{km})$ the $k$-th row of the transition matrix, and $\mathbf{z}_{s:t} = (\mathbf{z}_s, \ldots, \mathbf{z}_t)'$, $1 \leq s \leq t \leq T$, denotes a subsequence of a given sequence of variables, $\mathbf{z}_t$, $t = 1, \ldots, T$.

### 3.2 Prior Elicitation

In a Bayesian framework we need to complete the description of the model by specifying the prior distributions of the parameters. Proper priors may be undesiderable because they require subjective input. Thus, we assume objective priors (see Robert [2001], Ch. 3, for an introduction to prior elicitation), which are priors that do not use subjective input (see Kass and Wasserman [1996]) and yield posteriors with good frequentist properties, such as the second-order correct coverage for the intervals (see Wasserman [2000]). Unfortunately, the use of improper priors as objective priors in a context of (dynamic) mixture models may yield improper posterior distributions. This may happen with a positive probability when data provide no information about the parameters of one of the components (regime) of the (dynamic) mixture.

We overcome the impropriety problem by considering the data-dependent prior approach suggested by Diebolt and Robert [1994]. It has been shown (see Wasserman [2000], Th. 6) that a posterior distribution based on a data-dependent prior is identical with the posterior based on a Jeffreys prior, if the likelihood function is replaced with a pseudo-likelihood function. The choice of the data-dependent prior and of the pseudo-likelihood is such that the posterior is well defined. Given that we can reject simulated allocation variable draws in that part of the likelihood corresponding corresponding to troublesome grouping of the data, it follows that the posterior is proper. We shall see, in the next section, that the pseudo-likelihood can be easily computed during the posterior simulation, by imposing some constraints on the simulation of the allocation variables.

Since the results of Wasserman [2000] also apply to the improper priors other than the Jeffreys prior, we follow Diebolt and Robert [1994] and consider a conjugate partially improper prior. Conjugate improper priors are numerically close to the Jeffreys prior, provide similar inferences and yield easier posterior simulations. We assume uniform prior distributions for all the autoregressive coefficients, the intercept and the precision parameters

$$\begin{aligned}
(\phi_1, \ldots, \phi_p) &\propto \mathbb{I}_{\mathbb{R}^p}(\phi_1, \ldots, \phi_p) \\
\nu_k &\propto \mathbb{I}_{\mathbb{R}}(\nu_k), \quad k = 1, \ldots, m \\
\sigma_k^2 &\propto \frac{1}{\sigma_k^2} \mathbb{I}_{\mathbb{R}_+}(\sigma_k^2), \quad k = 1, \ldots, m
\end{aligned}$$

and do not impose stationarity constraints for the autoregressive coefficients. Sufficient conditions for the second order stationarity of MS-ARMA models are given in Francq and Zakoian [2001]. In particular, the literature has devoted great attention to the elicitation of suitable noninformative priors for the autoregressive coefficients (see De Pooter et al. [2008] for a review) and the use of Jeffreys prior is controversial in this setting (see Robert [2001], Note 4.7.2). Sims [1988] and Sims and Uhlig [1991] advocate the use of flat priors, while Phillips [1991] finds that flat priors bias the inference towards stationarity and suggests instead the use of Jeffreys priors. Moreover, prediction is often much more sensitive than parameter inference to the choice of the priors. Koop et al. [1995] show that imposing stationarity constraints on autoregressive coefficients of an AR(1) model needs not lead to stabilization of the predictive variance as the forecast horizon increases. Finally, we note that our model could be extended up to include regime-dependent autoregressive coefficients, with stationary coefficients in at least one of the regimes. Ang and Bekaert [2002] and Holst et al. [1994] prove that such processes retain covariance stationarity as long as the unconditional autocorrelation is strictly less than one. This is guaranteed by appropriate mixing of the regimes. With constant transition probabilities, a sufficient condition is that the ergodic probability associated with the stationary regime is non-zero. These models capture possible variations in the stationarity of the variable of interest and has been found useful in applied economic time series analysis, for instance for modelling GDP (McCulloch and Tsay [1994]) and interest rates (Ang and Bekaert [2002]).

We assume standard conjugate prior distributions for the transition probabilities. These distributions are independent and identical Dirichlet distributions, one for each row of the transition matrix

$$(p_{k1}, \ldots, p_{km})' \sim \mathcal{D}(\delta_1, \ldots, \delta_m)$$

with $k = 1, \ldots, m$.

When estimating a MS model, which is a dynamic mixture model, one needs to deal with the identification issue arising from the invariance of the likelihood function and of the posterior distribution (which follows from the assumption of symmetric prior distributions) to permutations of the allocation variables. Many different ways to solve this problem are discussed, for example, in Frühwirth-Schnatter [2006]. We identify the regimes by imposing some constraints on the parameters, as it is standard in business cycle analysis. We consider the following identification constraints on the intercept: $\nu_1 < 0$ and $\nu_1 < \nu_2 < \ldots < \nu_m$, which allow us to interpret the first regime as the one associated with the recession phase. As an alternative, one could introduce the constraints on the volatility or on the transition probabilities. From a practical point of view, we find in our empirical applications that volatility ordering works as well as the intercept ordering constraint for the regime identification. The ordering on the transition probabilities is not strong enough for the data to identify the regimes.

### 3.3 Posterior Simulation

Samples from the joint posterior distribution of the parameters and the allocation variables are obtained by iterating a Gibbs sampling algorithm. The full conditional distributions of the Gibbs sampler are given in the following together with the sampling procedure for the posterior of the allocation variables (see also Krolzig [1997]).

Let us introduce the auxiliary variables $y_{0t} = y_t - \boldsymbol{\xi}_t'\boldsymbol{\nu}$ and define: $\boldsymbol{\nu}_{-k} = (\nu_1, \ldots, \nu_{k-1}, \nu_{k+1}, \ldots, \nu_m)'$ and $\boldsymbol{\sigma}_{-k} = (\sigma_1, \ldots, \sigma_{k-1}, \sigma_{k+1}, \ldots, \sigma_m)'$. The full conditional distribution of the regime-independent parameter $\boldsymbol{\phi}$ is normal with density function

$$
\begin{aligned}
& f(\boldsymbol{\phi}|\mathbf{y}_{1:T}, \boldsymbol{\xi}_{1:T}, \boldsymbol{\nu}, \boldsymbol{\sigma}, \mathbf{p}) \propto \exp\left\{-\frac{1}{2}\sum_{t=1}^{T}(y_{0t} - \mathbf{x}_{0t}'\boldsymbol{\phi})^2\gamma_t^{-1}\right\} \qquad (5) \\
& \propto \exp\left\{-\frac{1}{2}\boldsymbol{\phi}'\left(\sum_{t=1}^{T}\mathbf{x}_{0t}\gamma_t^{-1}\mathbf{x}_{0t}'\right)\boldsymbol{\phi} + \boldsymbol{\phi}'\left(\sum_{t=1}^{T}\mathbf{x}_{0t}\gamma_t^{-1}y_{0t}\right)\right\} \\
& \propto \mathcal{N}_p(\boldsymbol{\mu}_\phi, \Upsilon_\phi)
\end{aligned}
$$

where $\boldsymbol{\mu}_\phi = \Upsilon_\phi(\sum_{t=1}^{T}\mathbf{x}_{0t}\gamma_t^{-1}y_{0t})$ and $\Upsilon_\phi = \left(\sum_{t=1}^{T}\mathbf{x}_{0t}\gamma_t^{-1}\mathbf{x}_{0t}'\right)^{-1}$. The improper prior for $\boldsymbol{\phi}$ yields a proper posterior for all possible values of the allocation variables $\boldsymbol{\xi}_t$. Thus, inference on this part of the parameter vector does not suffer the impropriety problem.

The full conditional distributions of the intercept parameters $\nu_k$, $k = 1, \ldots, m$, are normal with density function

$$
\begin{aligned}
& f(\nu_k|\mathbf{y}_{1:T}, \boldsymbol{\xi}_{1:T}, \boldsymbol{\phi}, \boldsymbol{\nu}_{-k}, \boldsymbol{\sigma}, \mathbf{p}) \propto \exp\left\{-\frac{1}{2}\sum_{t\in\mathcal{T}_k}u_t^2\gamma_t^{-1}\right\} \qquad (6) \\
& \propto \exp\left\{-\frac{1}{2}\nu_k^2\left(\sum_{t\in\mathcal{T}_k}\gamma_t^{-1}\right) + \nu_k\left(\sum_{t\in\mathcal{T}_k}\gamma_t^{-1}y_{1t}\right)\right\} \\
& \propto \mathcal{N}(\mu_k, \omega_k^2)
\end{aligned}
$$

with $\mu_k = \omega_k^2(\sum_{t\in\mathcal{T}_k}y_{1t}\gamma_t^{-1}) = T_k^{-1}\sum_{t\in\mathcal{T}_k}y_{1t}$ and $\omega_k^2 = \left(\sum_{t\in\mathcal{T}_k}\gamma_t^{-1}\right)^{-1} = \sigma_k^2T_k^{-1}$, where we defined $\mathcal{T}_k = \{t \in \{1, \ldots, T\}|\xi_{kt} = 1\}$, $T_k = \sum_{t\in\mathcal{T}_k}\xi_{kt}$, and $y_{1t} = y_t - \mathbf{x}_{0t}'\boldsymbol{\phi}$. For the intercept parameters, since we assume improper priors, the posteriors are not always proper distributions. The posterior is not proper if $\omega_k^{-2} \leq 0$ or, equivalently, if there are no observations allocated to the $k$-th regimes (i.e., $\mathcal{T}_k$ is empty). It is possible to avoid this offensive grouping of the data by rejecting, at each iteration of the Gibbs sampler, the draws of the sequence of allocation variables, $\boldsymbol{\xi}_t$, $t = 1, \ldots, T$, that do not belong to the set $\mathcal{S}_\nu = \{\boldsymbol{\xi}_{1:T} | \sum_{t=1}^{T} \xi_{jt} \geq 1, \forall j = 1, \ldots, m\}$. We will show how to deal with this issue when presenting the simulation procedure for the allocation variables.

The full conditional distributions of the precision parameters, $\sigma_k^{-2}$, $k = 1, \ldots, m$, are gamma with density

$$
\begin{aligned}
& f(\sigma_k^{-2} | \mathbf{y}_{1:T}, \boldsymbol{\xi}_{1:T}, \boldsymbol{\phi}, \boldsymbol{\nu}, \boldsymbol{\sigma}_{-k}, \mathbf{p}) \propto \sigma_k^{-2} \prod_{t \in \mathcal{T}_k} (\sigma_k^2)^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2} \sum_{t \in \mathcal{T}_k} u_t^2 \gamma_t^{-1} \right\} \\
& \propto (\sigma_k^2)^{-\left(\frac{T_k}{2}+1\right)} \exp\left\{ -\frac{1}{2\sigma_k^2} \left( \sum_{t \in \mathcal{T}_k} u_{kt}^2 \right) \right\} \\
& \propto \mathcal{G}a(\alpha_k/2, \beta_k/2)
\end{aligned} \tag{7}
$$

where $u_{kt} = y_t - \mathbf{x}_{0t}'\boldsymbol{\phi} - \nu_k$, $\alpha_k = T_k$ and $\beta_k = \sum_{t \in \mathcal{T}_k} u_{kt}^2$. The posterior is well defined if $\alpha_k > 0$, that holds true if there are at least 2 observations allocated to the regime $k$, or equivalently, if $T_k > 1$. To have proper posterior distributions we merely omit the values of the latent vectors, $\boldsymbol{\xi}_t$, $t = 1, \ldots, T$, that create impropriety. That comes to restrict sampling of $\boldsymbol{\xi}_{1:T}$ to the set $\mathcal{S}_\sigma = \{\boldsymbol{\xi}_{1:T} | \sum_{t=1}^{T} \xi_{jt} \geq 2, \forall j = 1, \ldots, m\}$. We shall account for this constraint when discussing generation of the allocation variables.

The full conditional distribution of the $k$-th row, $k = 1, \ldots, m$, of the transition matrix is

$$
\begin{aligned}
& f(\mathbf{p}_{k\cdot} | \mathbf{y}_{1:T}, \boldsymbol{\xi}_{1:T}, \boldsymbol{\phi}, \boldsymbol{\nu}, \boldsymbol{\sigma}, \mathbf{p}_{-k}) \propto \prod_{j=1}^{m} p_{kj}^{\delta_j} \prod_{t=1}^{T} \prod_{j=1}^{m} p_{kj}^{\xi_{jt}\xi_{kt}} \\
& \propto \mathcal{D}(\delta_1 + N_{k1}, \ldots, \delta_m + N_{km})
\end{aligned} \tag{8}
$$

where $\mathbf{p}_{-k} = (\mathbf{p}_{1\cdot}, \ldots, \mathbf{p}_{k-1\cdot}, \mathbf{p}_{k+1\cdot}, \ldots, \mathbf{p}_{m\cdot})'$ and

$$
N_{kj} = \sum_{t=1}^{T} \mathbb{I}_{\{j\}}(s_t) \mathbb{I}_{\{k\}}(s_{t-1})
$$

counts the number of transitions of the chain from the state $k$ to the state $j$.

In Krolzig [1997] the multi-move Gibbs sampler (see Carter and Kohn [1994] and Shephard [1994]) is presented for Markov-switching vector autoregressive models as an alternative to the single-move Gibbs sampler given, for example, in Albert and Chib [1993].

The multi-move procedure, also known as forward-filtering backward sampling (FFBS) algorithm, is particularly useful in our context because the Gibbs sampler makes use of two relevant quantities, the filtering and the smoothing probabilities, that can be used for turning point analysis.

The filtering probability at time $t$, $t = 1, \dots, T$, is determined by iterating the prediction step

$$p(\boldsymbol{\xi}_t = \boldsymbol{\iota}_j | \mathbf{y}_{1:t-1}) = \sum_{i=1}^{m} p(\boldsymbol{\xi}_t = \boldsymbol{\iota}_j | \boldsymbol{\xi}_{t-1} = \boldsymbol{\iota}_i) p(\boldsymbol{\xi}_{t-1} = \boldsymbol{\iota}_i | \mathbf{y}_{1:t-1}) \tag{9}$$

and the updating step

$$p(\boldsymbol{\xi}_t | \mathbf{y}_{1:t}) \propto p(\boldsymbol{\xi}_t | \mathbf{y}_{1:t-1}) p(y_t | \mathbf{y}_{t-1-p:t-1}, \boldsymbol{\xi}_t) \tag{10}$$

where $p(\boldsymbol{\xi}_t = \boldsymbol{\iota}_j | \boldsymbol{\xi}_{t-1} = \boldsymbol{\iota}_i) = p_{ij}$, with $\boldsymbol{\iota}_m$ the $m$-th column of the identity matrix and $p(y_t | \mathbf{y}_{t-p-1:t-1}, \boldsymbol{\xi}_t)$ the conditional distribution of the variable $y_t$ from a MSIH($m$)-AR($p$).

We shall notice that the prediction step can be used at time $t$ to find the predictive density of $\boldsymbol{\xi}_{t+1}$

$$p(\boldsymbol{\xi}_{t+1} | \mathbf{y}_{1:t}) \propto P' \, p(\boldsymbol{\xi}_t | \mathbf{y}_{1:t}) \tag{11}$$

and the one of $y_{t+1}$

$$p(y_{t+1} | \mathbf{y}_{1:t}) = \sum_{i=1}^{m} p(\boldsymbol{\xi}_{t+1} = \boldsymbol{\iota}_i | \mathbf{y}_{1:t}) p(y_{t+1} | \mathbf{y}_{t+1-p:t}, \boldsymbol{\xi}_{t+1}) \tag{12}$$

which, for a Gaussian MS-AR process, is a discrete mixture of normal distributions.

The smoothing probabilities given by

$$p(\boldsymbol{\xi}_t = \boldsymbol{\iota}_j | \mathbf{y}_{1:T}) \propto \sum_{i=1}^{m} p(\boldsymbol{\xi}_t = \boldsymbol{\iota}_j | \boldsymbol{\xi}_{t+1} = \boldsymbol{\iota}_i, \mathbf{y}_{1:t}) p(\boldsymbol{\xi}_{t+1} = \boldsymbol{\iota}_i | \mathbf{y}_{1:T}) \tag{13}$$

are evaluated recursively and backward in time for $t = T, T-1, \dots, 1$. These quantities are the posterior probabilities of the observation $y_t$ to be in one of the $m$ regimes at time $t$, given all the information available from the full sample of data. The conditional distribution $p(\boldsymbol{\xi}_t | \boldsymbol{\xi}_{t+1}, \mathbf{y}_{1:t})$, that is the building block of the smoothing probability formula, is used in the FFBS algorithm to sample the allocation variables from their joint posterior distribution sequentially and backward in time for $t = T, T-1, \dots, 1$. See Frühwirth-Schnatter [2006], ch. 11-13, for further details.

As discussed in previous sections, when using data-dependent priors the generation of the allocation variables should omit draws that yield to impropriety of the posterior. In our prior settings, the set of non-troublesome grouping is $\mathcal{S} = \mathcal{S}_\nu \cap \mathcal{S}_\sigma = \mathcal{S}_\sigma$. Thus, each time the set of allocation variables $\boldsymbol{\xi}_{1:T}$, does not assign at least two observations to each component of the dynamic mixture, the entire set $\boldsymbol{\xi}_{1:T}$, is rejected and a new set is drawn until a proper set is obtained.

The smoothing probabilities are usually employed also to detect the turning points. In this paper, we will not consider the cycle generated by the smoothing probabilities and instead applied a non-parametric approach (see the next section) to extract the turning points from the forecasting values of $y_{t+h}$.

# 4 Combining Linear and Non-linear Models

In this section we describe the rules used for combining the forecasts from linear (the AR) and non-linear (MS-AR) models and for predicting the turning points of the business cycle. We propose combining the models through use of two alternative schemes. The first one is a Bayesian Model Averaging (BMA) procedure based on the forecasting performance for the variable of interest. The second one is based on the performance of the models in terms of turning point forecasts.

The BMA procedure gives a combined predictive density $p(\tilde{y}_{t+1}|\mathbf{y}_{1:t})$ for the value $y_{t+1}$ using the information available up to time $t$, $t = 1, \ldots, T$, from a set of models $M_j$, $j = 1, ..., M$:

$$p(\tilde{y}_{t+1}|\mathbf{y}_{1:t}) = \sum_{j=1}^{M} w_{j,t+1} p(\tilde{y}_{t+1}|\mathbf{y}_{1:t}, M_j) \tag{14}$$

where $w_{j,t+1}$ is the $(0, 1)$-valued weight given to model $M_j$ computed at time $t$ and $p(\tilde{y}_{t+1}|\mathbf{y}_{1:t}, M_j)$ is the predictive density of $\tilde{y}_{t+1}$ conditional on model $M_j$, with $j =$ AR, MS-AR, and on the information available up to time $t$. It should be noticed that the point forecast, $\tilde{y}_{t+1}$, from the combined predictive density is a linear combination of the individual point forecasts $\tilde{y}_{j,t+1}$, computed as the median of the densities $p(\tilde{y}_{j,t+1}|\mathbf{y}_{1:t}, M_j)$, $j =$ AR, MS-AR.

To assess the forecast accuracy of each model, we follow recent studies in using the predictive likelihood of the model. Sources such as Geweke [1999] and Geweke and Whiteman [2006] emphasize the close relationship between the predictive likelihood and marginal likelihood, previously used in BMA and, more generally, as Bayesian evaluation criterion. As stated in Geweke (1999, p.15), "... the marginal likelihood summarizes the out-of-sample prediction record... as expressed in ... predictive likelihoods." See Bjørnland et al. [2009] and Hoogerheide et al. [2010] for similar recent applications and Terui and van Dijk [2002] for an alternative approach to forecast combination of linear and nonlinear time series models.

The cumulative predictive-likelihood at time $t+1$ associated to the $j$-th model is defined as

$$\eta_{j,t+1}^{PL} = \prod_{s=1}^{t+1} p(y_s|\mathbf{y}_{1:s-1}, M_j) \tag{15}$$

where $p(y_s|\mathbf{y}_{1:s-1}, M_j)$ is the (simulated) predictive density obtained from the model $j$ and evaluated at $y_s$. We build the weights for the $j$-th model, as

$$w_{j,t+1}^{PL} = \frac{\eta_{j,t}^{PL}}{\sum_{k=1}^{K} \eta_{j,t}^{PL}} \tag{16}$$

with $j =$ AR, MS-AR.

We also suggest combining the forecasts by applying some performance measures that are usually employed in the analysis of the turning points[1]. These statistics evaluate the ability of the AR and MS-AR to predict turning points with position and frequency similar to those of the turning points in a reference cycle. In this paper, we consider one of the most used measures, that is the concordance statistic for regular periodic behavior in the business cycles proposed by Harding and Pagan [2002]. This statistics is a non-parametric measure of the proportion of time during which two series, in our case the business cycle regimes, are in the same state. If the series take value of 1 in a expansion phase and 0 in a contraction phase, then the concordance measure ranges between 0 and 1, with 0 representing perfectly counter-cyclical switches, and 1 perfectly synchronous shifts. Obviously, for two regimes described by random walks, the measure will be 0.5 in the limit. Our combination approach could be extended up to include other statistics, such as the cumulative movements, the actual cumulative movements and the excess cumulated movements, suggested by Harding and Pagan [2002], to capture different characteristics of the cycle estimated with different models.

The turning point forecasts for the variable of interest, $x_t$, has been generated by the Bry and Boschan [1971] (BB) rule, that identifies a downward turn (or peak) at time $t$ if $x_{t-K} < x_t, \ldots, x_{t-1} < x_t$ and $x_t > x_{t+1}, \ldots, x_t > x_{t+K}$ and a upward turn (or trough) at time $t$ if $x_{t-K} > x_t, \ldots, x_{t-1} > x_t$ and $x_t < x_t, \ldots, x_t < x_{t+K}$. Similarly, we define a non-downward turn at time $t$ if $x_{t-K} < x_t, \ldots, x_{t-1} < x_t$ and $x_t < x_{t+1}, \ldots, x_t < x_{t+K}$ and a non-upward turn at time $t$ if $x_{t-K} > x_t, \ldots, x_{t-1} > x_t$ and $x_t > x_{t+1}, \ldots, x_t > x_{t+K}$. The parameter $K$ reduces the number of false signals. These definitions are standard in business cycle analysis (see for example Chauvet and Piger [2008]) and are also used (with

[1]See Clements and Harvey [2011] for a more general analysis on combinations of probability forecasts that are not restricted to be 0 or 1.

some adjustments) by the NBER institute for building the reference cycle for the US.

In the following we apply an approximation of the BB rule and use only downward, $D_t(K)$, and upward, $U_t(K)$, turn signals, that are

$$D_t(K) = \prod_{k=1}^{K} \mathbb{I}_{[x_{t-k},+\infty)}(x_t)\mathbb{I}_{[x_{t+k},+\infty)}(x_t) \tag{17}$$

$$U_t(K) = \prod_{k=1}^{K} \mathbb{I}_{(-\infty,x_{t-k}]}(x_t)\mathbb{I}_{(-\infty,x_{t+k}))}(x_t) \tag{18}$$

respectively. Our analysis can be extended to include modifications of the BB rule (see for example Mönch and Uhlig [2005]), which account for asymmetries and time-varying duration across business cycle phases.

We set $x_s = y_s$, $s = t-K,\ldots,t+K$, that is the actual industrial production growth rates, and get an indicator variable

$$z_{R,t} = z_{R,t-1}(1 - D_t(K)) + (1 - z_{R,t-1})U_t(K)$$

that is equal to 1 in the expansion phases and 0 in the recession phases. In our applications we consider a signal, $z_{R,t}$, $t = 1,\ldots,T$, generated with $K = 5$ and assume $z_{R,0}$ is given.

In the turning point prediction exercise we follow Canova and Ciccarelli [2004] and use the full predictive densities. More specifically we set $K = 1$ as in Canova and Ciccarelli [2004] and calculate the expected value of $D_t(K)$. We use the MCMC approximation of the predictive densities $p(y_{t+1}|\mathbf{y}_{1:t}, M_j)$, $j =$ AR, MS-AR, to evaluate the following downward turn probabilities

$$\begin{aligned} P_{j,t}^{(D)} &= \int_{-\infty}^{\infty} D_t(1)p(\tilde{y}_{t+1}|\mathbf{y}_{1:t}, M_j)d\tilde{y}_{t+1} \\ &= \int_{-\infty}^{\infty} \mathbb{I}_{[y_{t-1},+\infty)}(y_t)\mathbb{I}_{[\tilde{y}_{t+1},+\infty)}(y_t)p(\tilde{y}_{t+1}|\mathbf{y}_{1:t}, M_j)d\tilde{y}_{t+1} \end{aligned} \tag{19}$$

The combined predictive density in Eq. (14) is used to find a downward turn probability, that is denoted with $P_{BMA,t}$. The upward turn probabilities $P_{j,t}^{(U)}$, $j =$ AR, MS-AR, BMA, are defined similarly. Under the assumption of symmetric loss function, minimization of the expected loss leads to predict a peak at time $t$ if $P_{j,t}^{D} > 0.5$ and a trough if $P_{j,t}^{U} > 0.5$. The resulting cycle is:

$$z_{j,t} = z_{j,t-1}\mathbb{I}_{[0,0.5]}(P_{j,t}^{D}) + (1 - z_{j,t-1})\mathbb{I}_{(0.5,1]}(P_{j,t}^{U})$$

We stress that, in the proposed method, the prediction of a turning point at time $t$ needs the predictive density for the variable of interest $y_{t+k}$, $k = 1, \ldots, K$, from each model. As a consequence, our turning point detection strategy, for the MS-AR model, does not consider the results implied by the smoothed posterior probabilities as proposed in the literature (see for example Krolzig [2004]), but uses the predictive density of the observable variable integrated with respect to the latent Markov-switching variable. The advantage in using the marginal predictive, instead of the hidden state smoothed probabilities, is that the forecast will include all the information that is contained in both the observable variable and the latent states predictive densities.

We evaluate turning point forecasting ability of the different models by the concordance statistics given by

$$\eta_{j,t+1}^{CS} = \sum_{s=1}^{t+1} ((z_{j,s} z_{R,s}) - (1 - z_{j,s})(1 - z_{R,s})) \tag{20}$$

Although the concordance statistics could be used to compute BMA weights similarly to Eq. (16) and to combine the predictive densities, we follow an alternative route and use it to combine the phase indicator from the different models. The phase indicator variable that results from the combination must be a binary variable. Therefore, we propose combining the phased indicators by using weights that take value 0 or 1. In fact, for the concordance statistics, we adopt a model selection approach, which can be viewed as a very special case of model averaging. The model with the highest concordance with the reference cycle has a weight of 1, and the other models have null weights. In formula we have

$$w_{j,t+1}^{CS} = \mathbb{I}_{\{k_t^*\}}(j) \tag{21}$$

where $k_t^* = \arg\max\{\eta_{j,t}^{CS}, j = \text{AR}, \text{MS-AR}\}$.

# 5 Empirical Results

## 5.1 Data and Reference Cycle

In our study we consider the Industrial Production Index (IPI) from OECD at a monthly frequency for the United States (US), from February 1949 to January 2011, and for the Euro Area (EU), from January 1971 to January 2011. Data for both US and EU economies are seasonally and working day adjusted. We employ revised data from the April 2011 vintage, see Hamilton [2010] and Nalewaik [2011] for business cycle analysis using real-time data. In order to obtain the IPI at the Euro zone level a back-recalculation has been performed (see Anas et al. [2007a,b] and Caporin and Sartore [2006] for details). Since Phillips-Perron and Dickey-Fuller stationarity tests point out the non-stationarity of the IPI, we considered in our analysis the log-changes of the IPI index. The resulting series (see Fig. 1) are then used to detect and forecast the turning points.

Fig. 1 shows the reference cycle used in our analysis. The cycle is obtained by applying a BB rule to the US and EU IPI series. For comparative purposes, we show for the US economy the NBER official turning points, which are obtained by applying the BB rule with some adjustments on the whole series. The application of our rule allows for detection of the following contraction phases (from peak to trough) for the US economy since 1980M01:

- 1980 recession (1982M04-1982M12) which is within the NBER references dates;
- 1990 recession (1989M08-1991M01) which is not within the NBER references dates;
- internet bubble burst and 9/11 dates (2000M09-2002M02) which is within the NBER reference dates;
- short contraction (2002M11-2002M12) which is not within the NBER dates;
- sluggish recovery of the US economy and EU industrial recession. This made Greenspan and FED to keep rates very low (2003M03-2003M08);
- the 2007-2009 recession (2007M09-2009M08) which is within the NBER reference dates.

Following the results of the BB algorithm, the Euro area has experienced the follow contraction phases since January 1980M01:

- the second oil shock and US double dip recession (1980M09-1984M07);
- the 1986-87 recession (1986M06-1987M04);
- the 1992-94 recession (1992M05-1994M04);
- the Asian-crises related recession (1998M12-1999M07);
- the 2001 and 2003 industrial recessions (2001M09-2006M05);
- the 2007-09 recession (2008M09-2009M07).

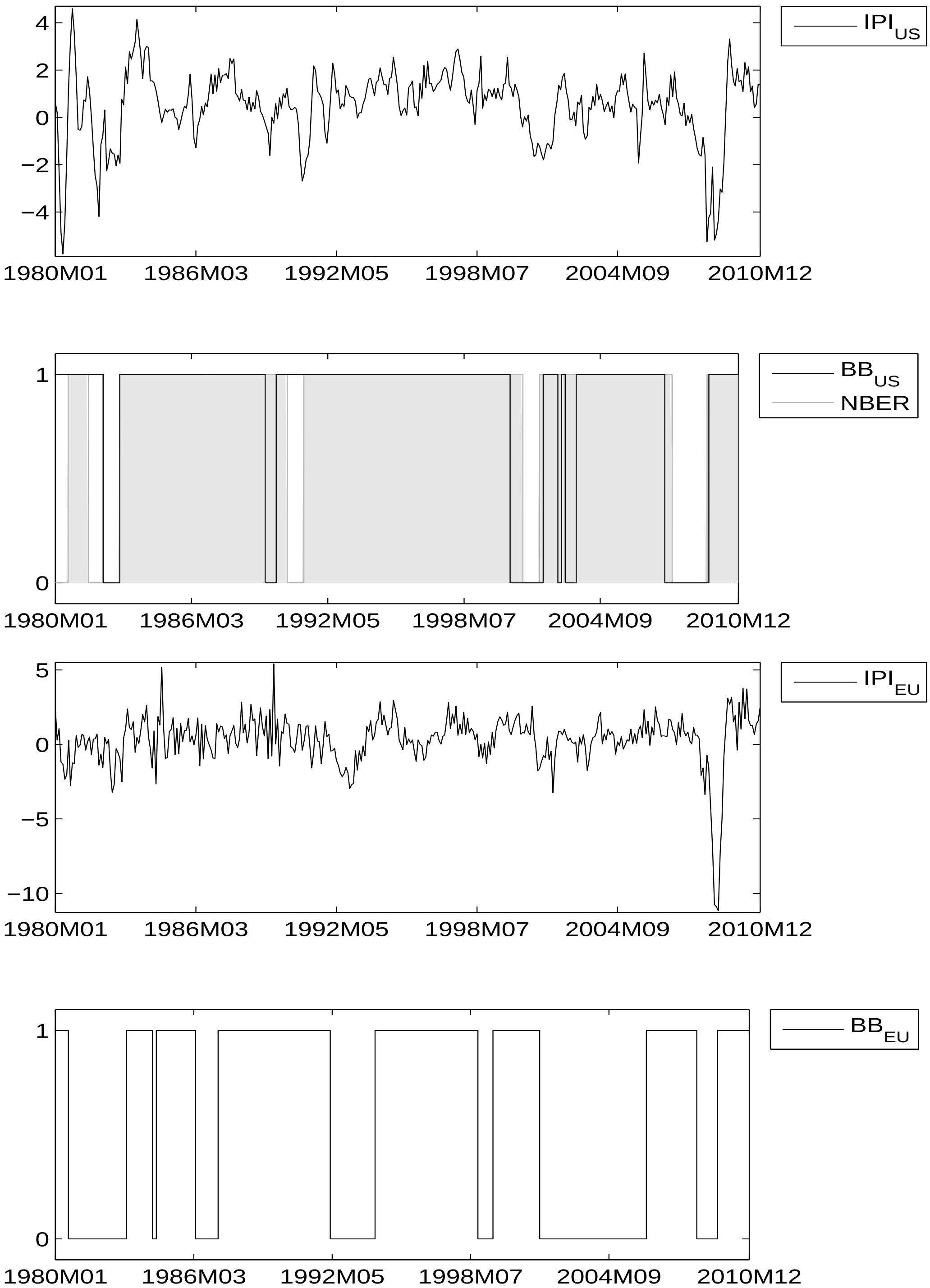


Figure 1: First and third chart: log-changes in the Industrial Production Index (IPI) for US and EU at monthly frequency for the period: January 1980 to January 2010. Second and fourth chart: the reference cycles (BB) for US and EU. Second chart: the NBER reference cycle (black).

| Country<br>Sample Period | US<br>1949M1-2011M1 | | | EU<br>1971M1-2011M1 | | |
|---|---|---|---|---|---|---|
| $\theta$ | $\hat{\theta}_T$ | $q_{0.05}$ | $q_{0.95}$ | $\hat{\theta}_T$ | $q_{0.05}$ | $q_{0.95}$ |
| $\nu$ | 0.111 | 0.042 | 0.181 | 0.074 | -0.039 | 0.189 |
| $\phi_1$ | 1.162 | 1.083 | 1.242 | 0.651 | 0.560 | 0.743 |
| $\phi_2$ | -0.105 | -0.222 | 0.012 | 0.339 | 0.237 | 0.441 |
| $\phi_3$ | -0.692 | -0.807 | 0.576 | -0.415 | -0.517 | -0.313 |
| $\phi_4$ | 0.795 | 0.670 | 0.920 | 0.187 | 0.096 | 0.278 |
| $\phi_5$ | -0.281 | -0.405 | 0.156 | | | |
| $\phi_6$ | -0.326 | -0.441 | 0.211 | | | |
| $\phi_7$ | 0.459 | 0.343 | 0.575 | | | |
| $\phi_8$ | -0.165 | -0.240 | 0.089 | | | |
| $\lvert\boldsymbol{\phi}\rvert$ | 0.848 | 0.808 | 0.889 | 0.763 | 0.702 | 0.825 |
| $\sigma$ | 3.891 | 1.124 | 6.634 | 2.357 | 1.379 | 6.011 |

Table 1: Estimated parameters of the AR($p$) model for the log-change of the US (with $p = 8$) and EU (with $p = 4$) Industrial Production Indexes. For each country: parameter estimates (first column) and the 0.05 and 0.95 quantiles (second and third columns).

### 5.2 Estimation and Forecasting

In the following we show the results of the sequential estimation and forecast of the AR and MS-AR models. The estimation results are based on 10,000 Gibbs iterations. The number of iterations has been chosen on the basis of both a graphical inspection of the Markov Chain Monte Carlo averages and on the application of the convergence diagnostic (CD) statistics proposed in Geweke [1992]. An initial set of 5,000 samples has been discarded to loose the dependence on the initial conditions of the sampler and the remaining samples were thinned down by a factor of 10 to have reasonably less-dependent posterior samples.

Table 1 shows the estimation results for the AR($p$) based on the full sample. We use the Bayesian information criteria for selecting the order of the autoregressive processes and find that for the US IPI log-changes an AR(8) should be used while an AR(4) should be considered for modelling the Euro area business cycle. For both of the cycles the AR($p$) has a positive intercept value that is statistically close to 0.1, which underestimates the mean value of the IPI log-changes during an expansion phase and overestimate it during a recession phases. The HPD region of the posterior distribution of the sum, in absolute value, of the autoregressive coefficients ($\lvert\boldsymbol{\phi}\rvert$ in Tab. 1) is the stationary region of the model The HPD region for the volatility is $(1.124, 6.634)$ for the US and $(1.379, 6.011)$ for the EU which are quite high and tend to overestimate volatility during the normal growth and the expansion periods.

| Country<br>Sample Period | US<br>1949M1-2011M1 | | | EU<br>1971M1-2011M1 | | |
|---|---|---|---|---|---|---|
| $\theta$ | $\hat{\theta}_T$ | $q_{0.05}$ | $q_{0.95}$ | $\hat{\theta}_T$ | $q_{0.05}$ | $q_{0.95}$ |
| $\nu_1$ | -2.436 | -5.868 | -0.539 | -1.981 | -4.225 | -0.423 |
| $\nu_2$ | -0.524 | -1.542 | 0.156 | -0.152 | -1.145 | 0.335 |
| $\nu_3$ | 0.132 | -0.290 | 0.511 | 0.482 | -0.040 | 1.433 |
| $\nu_4$ | 1.180 | 0.121 | 3.410 | 2.015 | 0.435 | 4.771 |
| $\sigma_1$ | 2.783 | 7.743 | 1.350 | 4.051 | 9.395 | 1.999 |
| $\sigma_2$ | 1.567 | 4.085 | 0.725 | 1.337 | 5.735 | 0.800 |
| $\sigma_3$ | 0.552 | 2.496 | 0.358 | 1.450 | 6.356 | 0.794 |
| $\sigma_4$ | 0.947 | 4.493 | 0.424 | 3.354 | 7.116 | 1.524 |
| $\phi_1$ | 0.935 | 0.650 | 1.204 | 0.555 | 0.330 | 0.774 |
| $\phi_2$ | 0.050 | -0.311 | 0.404 | 0.331 | 0.099 | 0.562 |
| $\phi_3$ | -0.473 | -0.815 | -0.135 | -0.374 | -0.610 | -0.140 |
| $\phi_4$ | 0.271 | 0.033 | 0.516 | 0.185 | -0.021 | 0.398 |
| $\lvert\boldsymbol{\phi}\rvert$ | 0.784 | 0.608 | 0.961 | 0.696 | 0.486 | 0.899 |
| $p_{11}$ | 0.641 | 0.429 | 0.832 | 0.709 | 0.519 | 0.870 |
| $p_{12}$ | 0.186 | 0.039 | 0.389 | 0.130 | 0.025 | 0.286 |
| $p_{13}$ | 0.082 | 0.004 | 0.240 | 0.079 | 0.004 | 0.223 |
| $p_{14}$ | 0.089 | 0.005 | 0.248 | 0.080 | 0.005 | 0.217 |
| $p_{21}$ | 0.041 | 0.002 | 0.136 | 0.032 | 0.001 | 0.130 |
| $p_{22}$ | 0.675 | 0.468 | 0.862 | 0.841 | 0.544 | 0.981 |
| $p_{23}$ | 0.165 | 0.031 | 0.359 | 0.090 | 0.005 | 0.281 |
| $p_{24}$ | 0.116 | 0.004 | 0.301 | 0.034 | 0.000 | 0.156 |
| $p_{31}$ | 0.014 | 0.000 | 0.048 | 0.051 | 0.000 | 0.207 |
| $p_{32}$ | 0.053 | 0.008 | 0.202 | 0.135 | 0.008 | 0.400 |
| $p_{33}$ | 0.886 | 0.596 | 0.976 | 0.775 | 0.414 | 0.980 |
| $p_{34}$ | 0.046 | 0.000 | 0.212 | 0.037 | 0.000 | 0.159 |
| $p_{41}$ | 0.033 | 0.000 | 0.139 | 0.096 | 0.005 | 0.272 |
| $p_{42}$ | 0.060 | 0.001 | 0.214 | 0.111 | 0.007 | 0.293 |
| $p_{43}$ | 0.128 | 0.005 | 0.319 | 0.115 | 0.008 | 0.288 |
| $p_{44}$ | 0.777 | 0.515 | 0.977 | 0.676 | 0.441 | 0.868 |

Table 2: Estimated parameters of the MSIH(4)-AR(4) model for the log-change of the US and EU Industrial Production Indexes. For each country: parameter estimates (first column) and the 0.05 and 0.95 quantiles (second and third columns).

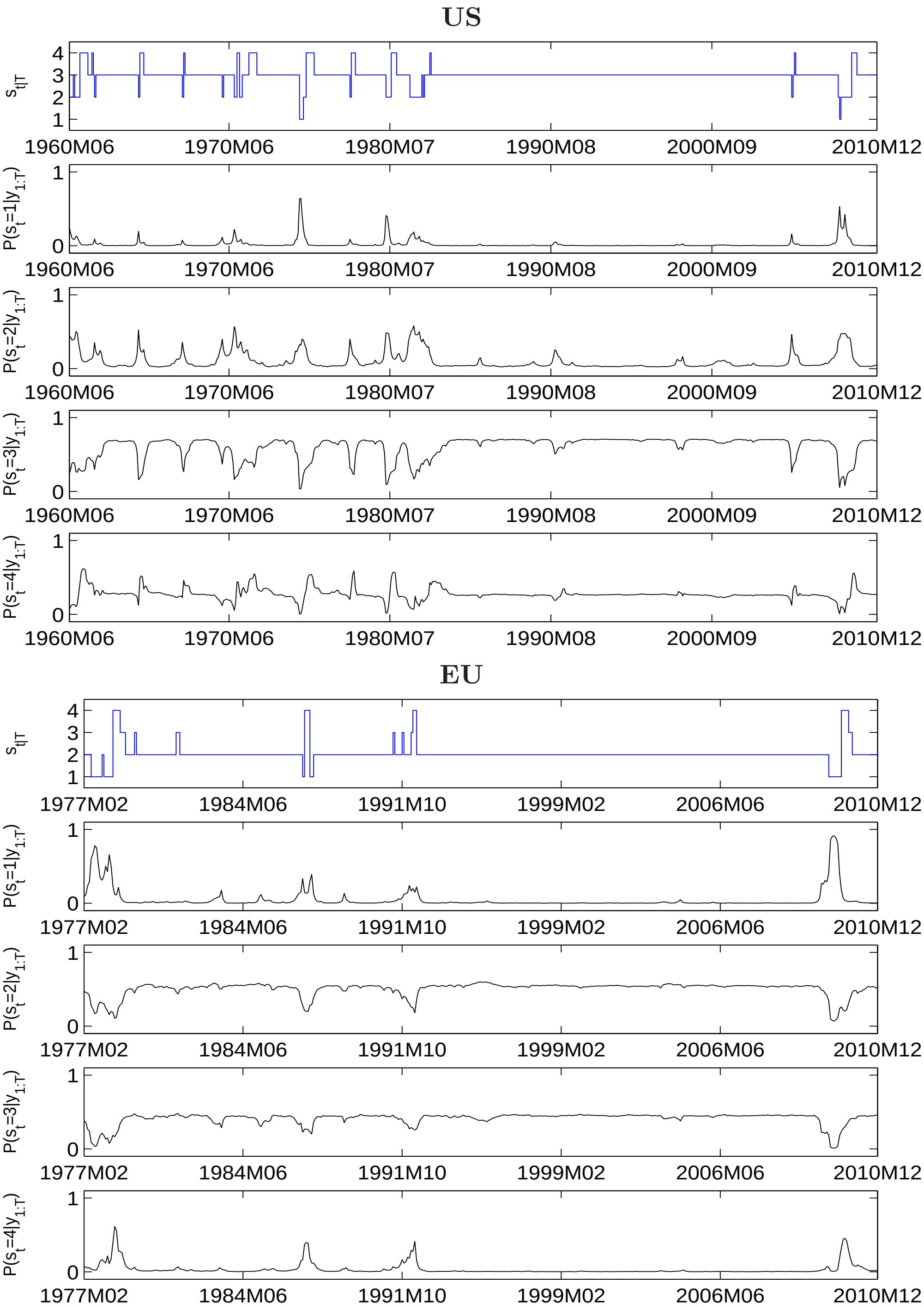


Figure 2: Hidden state estimates $s_{t|T}$ and smoothing probabilities $P(s_t|y_{1:T})$, for $t = 1, \ldots, T$, for US (upper panel) and EU (lower panel) data.

We compare the AR($p$) model with the MSIH($m$)-AR($p$) and as we expected the MSIH($m$)-AR($p$) are able to give a better description of the features of the cycles and to capture different phases in the IPI growth level and volatility. Tab. 2 shows the estimation results for the MSIH($m$)-AR($p$) based on the whole sample period. We consider here a flexible model by considering $p = 4$ lags as in Hamilton [1989] and Krolzig [2000] for the US gross domestic product and $m = 4$ regimes, extending the three-regimes model used in Krolzig [2000].

We find in our comparisons that the four-regimes model is necessary in order to capture the last recession. The interpretation of two of the four regimes will be similar to the one given in Krolzig [2000], i.e. normal growth and high growth, and two regimes are used to describe the recession phases. Thus in our model the fourth regime characterizes high-growth episodes, the third regime normal-growth phases, the second regime a normal slowdown in economic activity. The first regime may indicate strong-recession periods. We find evidence of the four regimes in both the US and the EU economies (see the first graph in both the US and the EU panels of Fig. 2). The graphs in the rows from two to four of Fig. 2 US and EU panels show the smoothing probabilities of the MSIH($m$)-AR($p$) model estimated on the full sample. The smoothing probabilities for the first regime, $P(s_t = 1|y_{1:T})$, show that some strong recession periods are present in the sample with a high probability. In particular, in the 1976 and 2009 crises for both the EU and US cycles there are some periods where the smoothing probabilities of the first regime are greater than the probabilities of the other regimes.

From Fig. 2 one can see that the regimes have different degrees of persistence. The analysis of the transition probabilities brings us to the following conclusions. The first regime is moderately persistent with transition probabilities $\hat{p}_{11} = 0.641$ for the US and $\hat{p}_{11} = 0.709$ for the EU (see Tab. 2). It is less persistent than the third regime (normal growth), which has estimated transition probabilities (see Tab. 2) $\hat{p}_{33} = 0.886$ for US and $\hat{p}_{33} = 0.775$ for EU. The second regime (normal recession) is less persistent than the other regimes, for US, with probability $\hat{p}_{22} = 0.675$ to stay in the regime, and more persistent, for EU, with transition $\hat{p}_{22} = 0.841$. The fourth regime is more persistent than the first regime, for the US, with probability $\hat{p}_{44} = 0.777$ to stay in the regime, while the opposite is the case for the EU, which has the probability of staying in a strong recession regime of $\hat{p}_{44} = 0.676$.

The four regimes have substantially different values for the intercept and scale parameters (see Tab. 2). The differences between the constant terms in the first and in the fourth regime are similar for the US and the EU, i.e. $(\hat{\nu}_4 - \hat{\nu}_1) = 3.616$ for the US and $(\hat{\nu}_4 - \hat{\nu}_1) = 3.996$ for the EU. The volatility gap between the first and fourth regimes is instead different in the two cycles: $\hat{\sigma}_4^2 - \hat{\sigma}_1^2 = -1.836$ for US and $\hat{\sigma}_4^2 - \hat{\sigma}_1^2 = -0.967$ for the EU. More generally the volatility of the EU cycle associated with regimes of strong recession and high growth is larger than the volatility of the US cycle. For both cycles the MS model results show that volatility significantly changes across the four regimes. For this reason, the use in this context of a AR model with constant volatility may be inappropriate. Accordingly, one could expect that the MS-AR models have superior forecasting ability than the AR models.

Fig. 3 shows the combination weights obtained from the sequential evaluation of the forecasting abilities of the different models for the US and the EU IPI log-changes. From the first and third chart in Fig. 3 it can be seen that the combination weights, $w^{PL}_{MS-AR,US}$ and $w^{PL}_{MS-AR,EU}$, increase in the last part of the sample, starting at September 2008. This corresponds to an increase in the forecasting ability, in terms of predictive likelihood, of the MS-AR with respect to the AR models. From our experiments we find that the good performance of the MS-AR models in the last part of the sample cannot be obtained with three regimes and that four-regime models are necessary to have an adequate description, in terms of expected growth-rate and volatility, of both the US and EU cycles during a strong recession phase.

The results for the performance abilities change if we consider the concordance with a reference cycle as a performance measure (see the combination weights $w^{CS}_{MS-AR,US}$ and $w^{CS}_{MS-AR,EU}$ in the second and fourth graph of Fig. 3). More specifically, for the US cycle (second chart in Fig. 3) the MS-AR model is superior to the AR model starting at the beginning of 1985. Conversely, the turning point forecast abilities of the MS-AR are worse than those of the AR model for the EU cycle, starting at the beginning of 1985. These results are all in line with the results in Clements and Krolzig [1998] about the time-varying performance of the MS models. MS models behave in a different way depending on the value of the regime present when the forecast performances are evaluated.

## 5.3 Sequential Turning Points Detection

Turning point prediction with different models (AR and MS-AR) and model combinations (using predictive likelihood and concordance statistics) are given in Fig. 4. Fig. 4 (charts 3 and 4) shows that the two combination strategies for the US cycle give two sequences of turning point forecasts that exhibit substantial differences. Charts seven and eight of the same figures show that the two strategies give similar turning points for the EU cycle.

In order to evaluate, at the end of the sample period $T$, the forecast abilities of the two

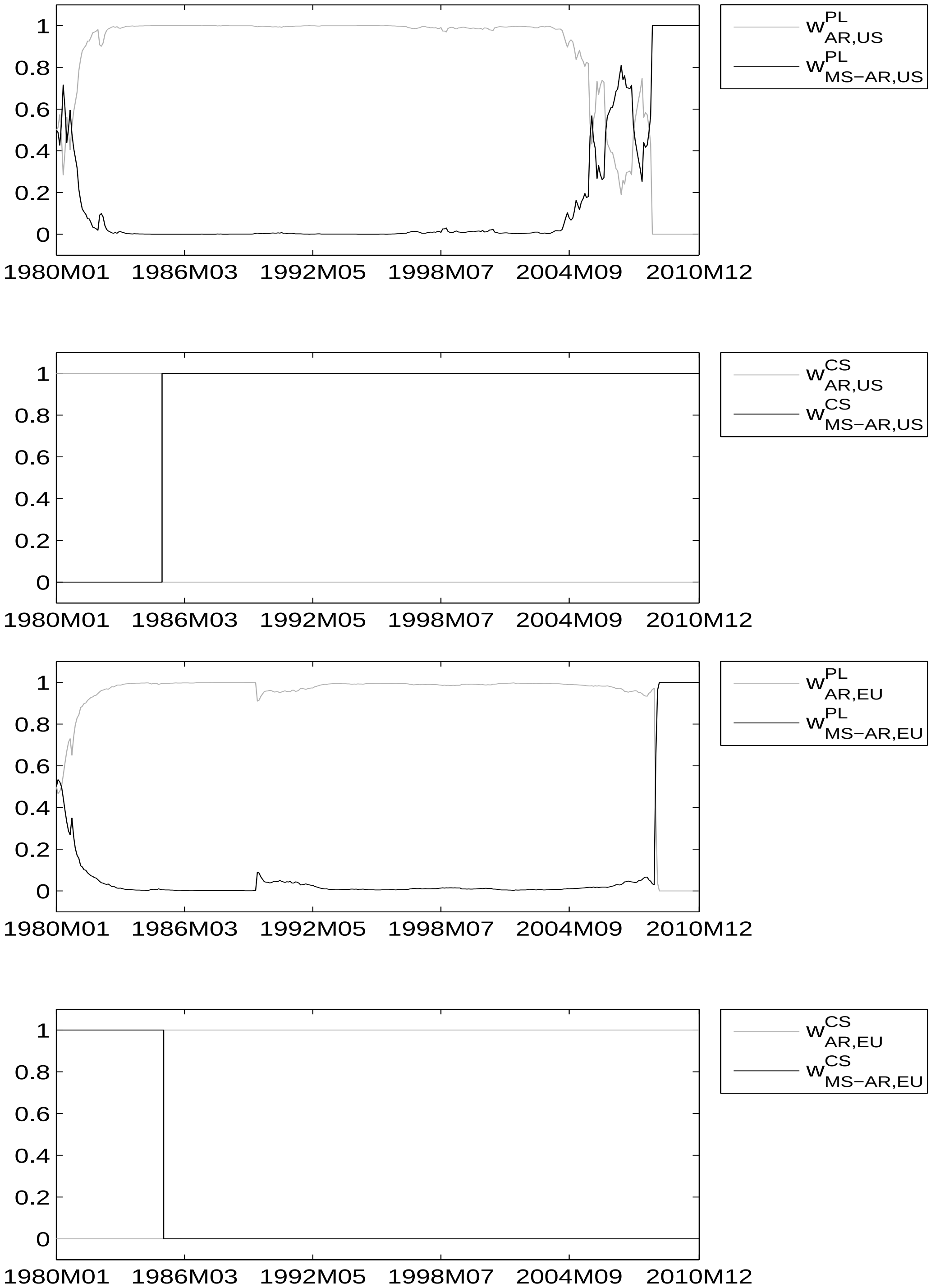


Figure 3: Combination weights for the AR($p$) and MSIH($m$)-AR($p$) forecasts by using predictive-likelihood (PL) and concordance statistics (CS) for US and EU data.

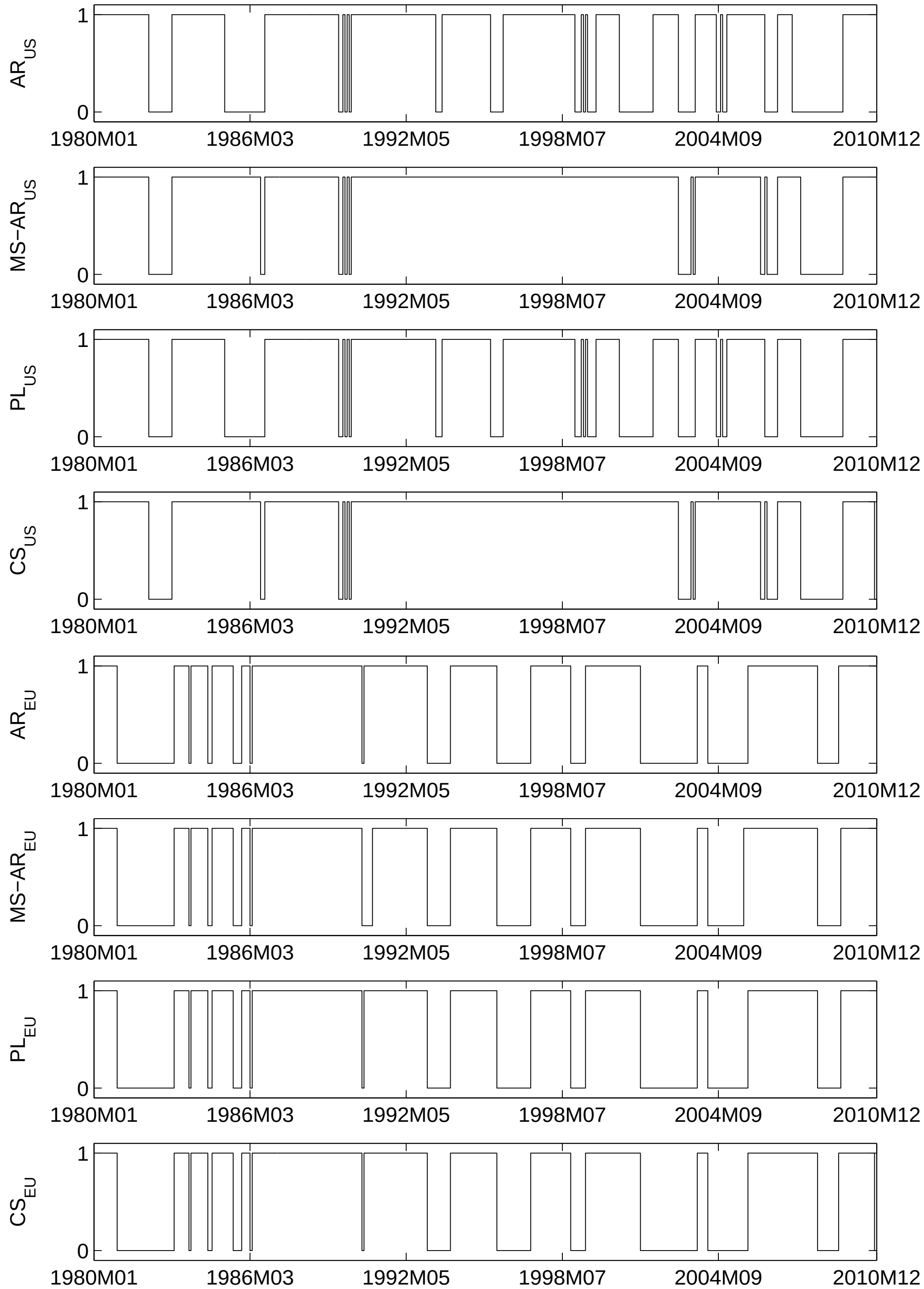


Figure 4: Turning point forecasts for US and EU IPI obtained from different models (AR($p$) and MSIH($m$)-AR($p$)) and their combinations based on the predictive likelihood (PL) and the concordance statistics (CS).

combination strategies we consider the Mean Square Prediction Error (MSPE)

$$MSPE = \frac{1}{T}\sum_{t=1}^{T}(y_t - \tilde{y}_{t+1})^2 \tag{22}$$

and the Logarithmic Score (LS)

$$LS = -\frac{1}{T}\sum_{t=1}^{T}\ln p(\tilde{y}_{t+1}|\mathbf{y}_{1:t}) \tag{23}$$

Tab. 3 shows that one of the two models performs better for both the US and EU, in terms of MSPE, than the two combination strategies. When considering the LS, then the forecast based on the concordance statistics that corresponds to the combination of the turning point indicators is the best strategy to use for the US cycle. For the EU cycle the forecast based on predictive likelihood performs better than the one based on concordance statistics. This leads to the conclusion that, for the EU it is better to combine first the growth-rate forecasts and then apply the BB rule for the detection of the turning points. Our findings are similar to Min and Zellner [1993]. They considered either the annual real GDP and real GNP of eighteen countries, 1974-87, and found that it is not always optimal to combine forecast when predicting the output growth rate.

| | AR | MS-AR | PL | CS |
|---|---|---|---|---|
| | US | | | |
| MSPE | 0.489 | 0.556 | 0.519 | 0.523 |
| LS | -1.200 | -1.144 | -1.209 | -1.121 |
| | EU | | | |
| MSPE | 1.323 | 1.299 | 1.299 | 1.331 |
| LS | -1.683 | -1.541 | -1.552 | -1.697 |

Table 3: Mean square prediction error (MSPE), Log-score (LS) for the AR($p$), MSIH($m$)-AR($p$) models and for the model combinations based on predictive likelihood (PL) and on the concordance statistics (CS).

## 6 Conclusion

We focus on the analysis the turning points of the business cycle and follow a Bayesian model averaging approach to combine their forecasts obtained from different prediction models. The new combination scheme relies upon non-parametric measures, i.e. concordance statistics, of the proportion of time during which the predicted and the reference chronologies, are in the same phase. We compare empirically our combination approach with a combination strategy based on the predictive likelihood, that detects the ability to predict the level of the cycle. In the comparison exercise we consider linear (AR) and nonlinear (MS-AR) models and follow a full Bayesian approach for inference and for both model estimation and combination, which accounts for both parameter and model uncertainty. In our finding, the predictive likelihood and the concordance statistics show that forecast abilities of the models change across different phases of the cycle. In our specific case, the two measures rank differently the predictive models. As a consequence, the performances of the different combination strategies are different. We also found that the results are cycle-specific and this suggest that both of them should be considered in the applications and the best combination evaluated for the problem at hand.

Finally, our analysis could be extended to include some generalisations of the model of Hamilton [1989] such as MS latent factor models (Kim and Nelson [1999]), MS models with time-varying transition probability (Sichel [1991], Watson [1994], Diebold and Rudebusch [1996], Durland and McCurdy [1994], and Filardo [1994]), stochastic duration models (Billio and Casarin [2010], Billio and Casarin [2011] and Chib and Dueker [2004]), and finally multivariate MS models (Diebold and Rudebusch [1996] and Krolzig [1997, 2004]). The degrees of freedom in the specification of the model set and of the concordance statistics makes the proposed methodology very appealing and suggests the application of our BMA approach to all the empirical analysis where forecasting of the turning points is a crucial issue.

## References


J. H. Albert and S. Chib. Bayes inference via Gibbs sampling of autoregressive time series subject to Markov mean and variance shifts. *Journal of Business and Economic Statistics*, 11:1–15, 1993.

J. Anas, M. Billio, L. Ferrara, and M. Lo Duca. Business Cycle Analysis with Multivariate Markov Switching Models. In G. L. Mazzi and G. Savio, editors, *Growth and Cycle in the Eurozone*, pages 249–260. Palgrave MacMillan, 2007a.

J. Anas, M. Billio, L. Ferrara, and M. Lo Duca. A turning point chronology for the Eurozone classical and growth cycle. In G. L. Mazzi and G. Savio, editors, *Growth and Cycle in the Eurozone*, pages 261–274. Palgrave MacMillan, 2007b.

J. Anas, M. Billio, L. Ferrara, and G. L. Mazzi. A System for Dating and Detecting Turning Points in the Euro Area. *The Manchester School*, 76:549–577, 2008.

A. Ang and G. Bekaert. Regime switches in interest rates. *Journal of Business and Economic Statistics*, 20:163–182, 2002.

M. Billio and R. Casarin. Identifying Business Cycle Turning Points with Sequential Monte Carlo Methods: an Online and Real-Time Application to the Euro Area. *Journal of Forecasting*, 29:145–167, 2010.

M. Billio and R. Casarin. Beta Autoregressive Transition Markov-switching Models for Business Cycle Analysis. *Studies in Nonlinear Dynamics and Econometrics*, 1:1–30, 2011.

H.C. Bjørnland, K. Gerdrup, A.S. Jore, C. Smith, and L. A. Thorsrud. Does forecast combination improve Norges Bank inflation forecasts? *Oxford Bulletin of Economics and Statistics*, Forthcoming, 2009.

G. Bry and C. Boschan. Cyclical Analysis of Time Series: Selected Procedures and Computer Programs. NBER Technical Paper 20, 1971.

F. Canova and M. Ciccarelli. Forecasting and turning point prediction in a bayesian panel var model. *Journal of Econometrics*, 120:327–359, 2004.

M. Caporin and D. Sartore. Methodological aspects of time series back-calculation. Working Paper, DSE, University of Venice, 2006.

C. K. Carter and R. Kohn. On Gibbs Sampling for State Space Models. *Biometrika*, 81(3): 541–553, 1994.

M. Chauvet and J. Piger. Comparison of the real-time performance of business cycle dating methods. *Journal of Business and Economic Statistics*, 26:42–49, 2008.

S. Chib and M. Dueker. Non-Markovian Regime Switching with Endogenous States and TimeVarying State Strengths. Working Paper 2004-030, Federal Reserve Bank of St. Louis, 2004.

M. P. Clements and H. M. Krolzig. A comparison of the forecast performances of Markov-switching and threshold autoregressive models of US GNP. *Econometrics Journal*, 1: C47–C75, 1998.

Michael P. Clements and David I. Harvey. Combining probability forecasts. *International Journal of Forecasting*, 27(2):208–223, 2011.

M. De Pooter, F. Ravazzolo, R. Segers, and H.K. van Dijk. Bayesian near-boundary analysis in basic macroeconomic time series models. *Advances in Econometrics*, 23:331–432, 2008.

F. X. Diebold and G. D. Rudebusch. Measuring business cycles: A modern perspective. *The Review of Economics and Statistics*, 78:67–77, 1996.

J. Diebolt and C. P. Robert. Estimation of finite mixture distributions through Bayesian sampling. *Journal of the Royal Statistical Society B*, 56:363–375, 1994.

J. M. Durland and T. H. McCurdy. Duration-depedent transition in Markov model of U.S. GNP growth. *Journal of Business and Economics Statistics*, 12:279–288, 1994.

A. F. Filardo. Business cycle phases and their transitional dynamics. *Journal of Business and Economics Statistics*, 12:299–308, 1994.

C. Francq and J. M. Zakoian. Stationarity of multivariate Markov-switching ARMA models. *Journal of Econometrics*, 102(2):339–364, 2001.

S. Frühwirth-Schnatter. *Mixture and Markov-swithing Models*. Springer, New York, 2006.

J. Geweke. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, editors, *Bayesian Statistics 4*, pages 169–193. Oxford University Press, Oxford, 1992.

J. Geweke. Using simulation methods for Bayesian econometric models: inference, development,and communication. *Econometric Reviews*, 18(1):1–73, 1999.

J. Geweke and C. Whiteman. Bayesian forecasting. In G. Elliot, C.W.J. Granger, and A.G. Timmermann, editors, *Handbook of Economic Forecasting*. North-Holland, 2006.

S. M. Goldfeld and R. E. Quandt. A Markov Model for Switching Regression. *Journal of Econometrics*, 1:3–16, 1973.

G. K. Grunwald, A. E. Raftery, and P. Guttorp. Time series of continuous proportions. *Journal of the Royal Statistical Society, Series B*, 55:103–116, 1993.

J. D. Hamilton. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, 57:357–384, 1989.

J. D. Hamilton. Calling recessions in real time. *International Journal of Forecasting*, forthcoming, 2010.

D. Harding and A. Pagan. Dissecting the Cycle: A Methodological Investigation. *Journal of Monetary Economics*, 49:365–381, 2002.

U. Holst, G. Lindgren, J. Holst, and M. Thuvesholmen. Recursive Estimation in Switching Autoregressions with a Markov Regime. *Journal of Time Series Analysis*, 15(5):489–506, 1994.

L. Hoogerheide, R. Kleijn, R. Ravazzolo, H. K. van Dijk, and M. Verbeek. Forecast Accuracy and Economic Gains from Bayesian Model Averaging using Time Varying Weights. *Journal of Forecasting*, 29(1-2):251–269, 2010.

M. Jochmann, G. Koop, and R.W. Strachan. Bayesian forecasting using stochastic search variable selection in a VAR subject to breaks. *International Journal of Forecasting*, 26(2): 326–347, 2010.

R. E. Kass and L. Wasserman. The selection of prior distributions by formal rules. *Journal of the American Statistical Association*, 91:1343–1370, 1996.

C. J. Kim and C. J. Murray. Permanent and Transitory Components of Recessions. *Empirical Economics*, 27:163–183, 2002.

C. J. Kim and C. R. Nelson. *State-Space Models with Regime Switching.* MIT press, Cambridge, 1999.

C. J. Kim and J. Piger. Common stochastic trends, common cycles, and asymmetry in economic fluctuations. Working paper, n. 681, International Finance Division, Federal Reserve Board, Semptember 2000, 2000.

G. Koop, J. Osiewalski, and M.F.J. Steel. Bayesian long-run predictioni in time series models. *Journal of Econometrics*, 69:61–80, 1995.

H.-M. Krolzig. *Markov Switching Vector Autoregressions. Modelling, Statistical Inference and Application to Business Cycle Analysis.* Springer, Berlin, 1997.

H.-M. Krolzig. Predicting Markov-Switching Vector Autoregressive Processes. Nuffield College Economics Working Papers, 2000-WP31, 2000.

H.-M. Krolzig. Constructing turning point chronologies with Markov-switching vector autoregressive models: the Euro-zone business cycle. In G. L. Mazzi and G. Savio, editors, *Modern Tools for Business Cycle Analysis*, pages 147–190. Eurostat, Luxembourg, 2004.

R. E. McCulloch and R. S. Tsay. Bayesian inference of trend- and difference-stationarity. *Econometric Theory*, 10(3-4):596–608, 1994.

C.K. Min and A. Zellner. Bayesian and non-bayesian methods for combining models and forecasts with applications to forecasting international growth rates. *Journal of Econometrics*, 56:89–118, 1993.

E. Mönch and H. Uhlig. Towards a monthly business cycle chronology for the euro area. *Journal of Business Cycle Measurement and Analysis*, 2:43–69, 2005.

J. J. Nalewaik. Estimating probabilities of recession in real time using GDP and GDI. *Journal of Money, Credit, and Banking*, forthcoming, 2011.

P.C.B. Phillips. Bayesian unit roots: De rebus prioribus semper est disputandum. *Journal of Applied Econometrics*, 6:435–474, 1991.

S. M. Potter. A Nonlinear Approach to U.S. GNP. *Journal of Applied Econometrics*, 10: 109–125, 1995.

C. P. Robert. *The Bayesian Choice, 2nd ed.* Springer Verlag, New York, 2001.

N. Shephard. Partial non-Gaussian state space. *Biometrika*, 81:115–131, 1994.

D. E. Sichel. Business cycle duration dependence: A parametric approach. *Review of Economics and Statistics*, 73:254–256, 1991.

C. Sims. Bayesian Skepticism on unit root econometrics. *Journal of Economic Dynamics and Control*, 12:463–474, 1988.

C. Sims and H. Uhlig. Understanding unit rooters: a helicopter tour. *Econometrica*, 59: 1591–1599, 1991.

J. H. Stock and M. W. Watson. Indicators for dating business cycles: Cross-history selection and comparisons. Working paper 2010, 2010.

M. Tanner and W. Wong. The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association*, 82:528–550, 1987.

N. Terui and H. K. van Dijk. Predictability in the Shape of the Term Structure of Interest Rates. *International Journal of Forecasting*, 18:421–438, 2002.

H. Tong. *Threshold Models in Non-Linear Time-Series Models.* Springer-Verlag, New York, 1983.

J. Vermaak, C. Andrieu, A. Doucet, and S. J. Godsil. Reversible jump Markov chain Monte Carlo strategies for Bayesian model selection in autoregressive processes. *Journal of Time Series Analysis*, 25:785–809, 2004.

L. Wasserman. Asymptotic inference for mixture models using data-dependent priors. *Journal of the American Statistical Association*, 62:159–180, 2000.

M. Watson. Business cycle durations and postwar stabilization of the U.S. economy. *American Economic Review*, 84:24–46, 1994.

A. Zellner, C. Hong, and C. Min. 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, 1991.