Model Averaging in Markov-Switching Models: Predicting National Recessions with Regional Data

Guérin, Pierre and Leiva-Leon, Danilo

17 October 2014

Model Averaging in Markov-Switching Models: Predicting National Recessions with Regional Data*

Pierre Guérin† Danilo Leiva-Leon‡

November 27, 2014

Abstract

This paper estimates and forecasts U.S. business cycle turning points with state-level data. The probabilities of recession are obtained from univariate and multivariate regime-switching models based on a pairwise combination of national and state-level data. We use two classes of combination schemes to summarize the information from these models: Bayesian Model Averaging and Dynamic Model Averaging. In addition, we suggest the use of combination schemes based on the past predictive ability of a given model to estimate regimes. Both simulation and empirical exercises underline the utility of such combination schemes. Moreover, our best specification provides timely updates of the U.S. business cycles. In particular, the estimated turning points from this specification largely precede the announcements of business cycle turning points from the NBER business cycle dating committee, and compare favorably with competing models.

Keywords: Markov-switching, Nowcasting, Forecasting, Business Cycles, Forecast combination.

JEL Classification Code: C53, E32, E37.

*We would like to thank Massimiliano Marcellino for helpful comments and suggestions on a previous version of this paper as well as Lutz Kilian, Michael Owyang, James Hamilton, Ana Galvão, Jeremy Piger and seminar participants at the Applied Time Series Econometrics Workshop held at the Federal Reserve Bank of St. Louis for helpful discussions. The views expressed in this paper are those of the authors. No responsibility for them should be attributed to the Bank of Canada.

†Bank of Canada, e-mail: pguerin@bank-banque-canada.ca

‡Bank of Canada, e-mail: leiv@bank-banque-canada.ca

1 Introduction

Assessing the current state of the economy represents a key input for policy makers and investors in their decision-making process. However, current economic conditions are typically subject to substantial uncertainty owing to the publication delay of macroeconomic variables and data revisions that become available long after the initial estimates have been released. Likewise, detecting business cycle turning points in real-time proves to be very challenging. As a result, economists have developed tools to provide early assessments of business cycle turning points. In particular, regime switching models have long been used to date and predict turning points. An important aspect of this type of models is that regime changes are endogenously estimated in a purely data driven way.

Since the seminal work of Hamilton (1989), a number of extensions to regime-switching models have been proposed to estimate turning points for the U.S. economy. In this context, dynamic factor models subject to regime changes are one of the most successful approaches. Relevant contributions include Kim (1994), Kim and Yooa (1995) and Kim and Nelson (1998). Moreover, Chauvet (1998) finds that this type of models perform well to date business cycle turning points in an out-of-sample experiment. Kholodilin and Yao (2005) use leading indicators in a dynamic factor model to predict turning points. Recent works have focused on analyzing the performance of regime switching models to forecast turning points using real-time data (Chauvet and Hamilton (2006) and Chauvet and Piger (2008)) and allowing for mixed frequency data (Camacho et al. (2012), Guérin and Marcellino (2013) and Camacho et al. (2014)).

Alternative approaches used to infer turning points rely on vector autoregressive (VAR) models with regime switching parameters. Relevant works include Hamilton and Perez-Quiros (1996) and Cakmakli et al. (2013) who use information on leading economic indexes to predict cycles for Gross National Product and Industrial Production, respectively. Nalewaik (2012) emphasizes the predictive content of gross domestic income (GDI) to forecast U.S. recessions in real-time. Finally, Hamilton (2011) provides a comprehensive survey of the literature on predicting turning points in real-time.

In a forecasting context, it is standard practise to rely on a combination of models to deal with model uncertainty and parameter instability. In fact, as discussed in Timmermann (2006), forecast combinations have frequently been found to produce better forecasts on average than methods based on the ex ante best individual forecasting model. An approach increasingly used in empirical studies is Bayesian Model Averaging (BMA), proposed by Raftery et al. (1998) and Hoeting et al. (1999) for linear models, which produces weights for each model that are constant over time.

Recently, Raftery et al. (2010) proposed a Dynamic Model Averaging (DMA) approach, where the models' weights are allowed to evolve over time. DMA has been applied to a variety of contexts. In detail, to Time-Varying Parameter (TVP) regression models to forecast inflation (Koop and Korobilis (2012), Chan et al. (2012) and Belmonte and Koop (2014)), the European Carbon market (Koop and Tole (2013)), and major monthly US macroeconomic variables using information from Google searches (Koop and Onorante (2013)). In a multivariate context, DMA has been applied to linear VAR models (Koop

(2014) and Koop and Onorante (2012)), large TVP-VAR models (Koop and Korobilis (2013)) to forecast inflation, real output and interest rates as well as factor models (Koop and Korobilis (2011)) to forecast growth and inflation in the U.K. An alternative approach for model combination is provided in Elliott and Timmermann (2005) that suggests to use weights that can vary according to Markov chains. Geweke and Amisano (2011) instead focus on constructing optimal weights by considering linear pools where the objective is to maximize the historical log of predictive score. Del Negro et al. (2013) provides a dynamic version of the linear pools approach.

Despite the extensive literature on model averaging to formulate continuous forecasts, little has been done regarding the study of averaging schemes in the context of discrete forecasts. To the best of our knowledge, there are very few related works in this research area. For example, Billio et al. (2012) compare the performance of combination schemes for linear and regime switching models, while Billio et al. (2013) propose a time-varying combination approach for multivariate predictive densities. Moreover, Berge (2013) compares model selection schemes based on boosting algorithms with a Bayesian model averaging weighting scheme for predicting U.S. recessions based on logistic regressions using a set of economic and financial indicators. He finds that the results are comparable across the different weighting schemes.

The purpose of this paper is to contribute to this literature by evaluating different model averaging approaches for Markov-switching models to nowcast and forecast business cycle turning points in real-time. Specifically, we compare the forecasting performance of averaging schemes based on constant weights (BMA) with those based on time-varying weights (DMA), adapting the DMA approach of Raftery et al. (2010) to (univariate and multivariate) Markov-switching models. Moreover, a key contribution of this paper is to propose another criterion to estimate the models' weights, which relies on the ability of each model to fit business cycle turning points. It therefore differs from the standard approach that is only based on the likelihood associated with each model to determine the weights.

In a Monte Carlo experiment, we study the relevance of these different weighting schemes. We do find evidence in favor of weighting schemes based on past predictive performance to classify regimes in that such combination schemes yield lower quadratic probability score (our evaluation metric to evaluate how well a model estimates regimes). This holds true for both BMA and DMA weighting schemes.

The empirical application concentrates on predicting U.S. national recessions using state-level data. In this context, it is natural to think of the best way to combine information from the different U.S. states to predict a national aggregate. Another reason for focusing on this application is to contribute to the literature on the relationship between regional and national macroeconomic developments. Owyang et al. (2005), Hamilton and Owyang (2012) and Leiva-Leon (2014) use state-level data to study the synchronization of business cycles across U.S. states, finding that despite the significant heterogeneity in cyclical fluctuations, states have become more synchronized since the mid-90s. Also, Owyang et al. (2014) use state-level data to forecast U.S. recessions from probit models, showing that enlarging a set of preselected national variables with state-level data on employment growth substantially improves nowcasts and short-term forecasts of the business cycle phases. We follow Owyang et al. (2014) in that we also use state-level employment data to predict national U.S. recessions. However, we focus on regime switching models rather than probit models that include the NBER datation of business cycle regimes as a dependent variable, which is problematic in a forecasting context given the substantial publication delay in the announcements of the NBER business cycle turning points.

The main results can be summarized as follows. First, we find that it is relevant to take into account the models' ability to estimate regimes when calculating models' weights if one is interested in regime classification. Indeed, our combination schemes based on the quadratic probability score typically outperform combination schemes based on the likelihood only. This is especially true in an out-of-sample context. Second, the use of regional data improves the forecasting performance of the models compared with models using exclusively national data. Third, the best forecasting model in the out-of-sample exercise outperforms the anxious index from the Survey of Professional Forecasters at short-forecasting horizons, which emphasizes the relevance of our framework. In addition, in a purely real-time environment, we also find that our best specifications provide timely estimates of the latest U.S. recession.

The paper is organized as follows. Section 2 describes the models we use, and Section 3 the different combination schemes we implement, and present the changes we make to the standard combination schemes. In Section 4, a small-sample Monte Carlo experiment is conducted to evaluate in a controlled experiment the combination schemes outlined in the previous section. Section 5 introduces the data, and details the results. Section 6 concludes.

2 Econometric Framework

2.1 Univariate model

We first consider a univariate regime-switching model defined as follows:

$$y_t = \mu_0^k + \mu_1^k S_t^k + \beta^k x_t^k + v_t^k \tag{1}$$

where y_t is a U.S. national indicator, x_t^k is total (non-farm) employment data for state k, both in growth rates, and v_t^k is the regression error term (it is assumed to be normally distributed, that is, $v_t^k \sim N(0, \sigma_k^2)$). S_t^k is a standard Markov-chain defined by the following constant transition probability:

$$p_{ij}^k = P(S_{t+1}^k = j | S_t^k = i), \tag{2}$$

$$\sum_{j=1}^{M} p_{ij}^k = 1 \forall i, j \epsilon \{1, ..., M\} \tag{3}$$

where M is the number of regimes.

Note that this specification differs from the baseline specification in Owyang et al. (2005) in that equation (1) mixes national data (i.e., y_t) with state-level data (i.e, the x_t^k's). In contrast, Owyang et al. (2005) estimate univariate regime-switching model on state-level data only to study the synchronization of economic activity across U.S. states. Moreover, Hamilton and Owyang (2012) examine the synchronization of U.S. states' business cycles using a panel data model under the assumption that a small number of clusters can explain the dynamics of U.S. states' business cycles. It is also worth mentioning the work of Owyang et al. (2014) that estimate a probit model to forecast U.S. recessions using a large number of covariates, including both national and state-level data. These authors then use Bayesian model averaging to select the most relevant predictors for forecasting U.S. recessions. Finally, a common feature of these works is to strive for parsimonious specifications to study business cycle dynamics, which is even more relevant in a forecasting context. This is guiding our modeling choice in equation (1) to study the relevance of state-level data to predict U.S. recessions.

In addition, we also use as a benchmark model a univariate regime-switching model with no exogenous predictor defined as:

$$y_t = \mu_0 + \mu_1 S_t + u_t \tag{4}$$

where $u_t \sim N(0, \sigma^2)$

2.2 Multivariate model

We then move on to consider a bivariate model where both the state-level data and the national data are stacked in the vector of dependent variables:

$$z_t = \Gamma(S_t^y, S_t^k) + \epsilon_t^k \tag{5}$$

where $z_t = (y_t, x_t^k)'$, and $\Gamma(S_t^y, S_t^k) = (\mu_0^y + \mu_1^y S_t^y, \mu_0^k + \mu_1^k(S_t^k))'$. y_t is the U.S. national indicator, and x_t^k is the total (non-farm) employment data for state k, both in growth rates. ϵ_t^k is normally distributed, and S_t^y and S_t^k are two independent Markov chains.

A few additional comments are required. First, we use a different Markov-chain (S_t^y and S_t^k) for each equation of the VAR, assuming that they are independently generated. This implies that regime changes at the national and state-level do not necessarily coincide, a feature that allows for the presence of the jobless recovery phenomenon observed in the U.S. employment data. Second, we do not include autoregressive dynamics in the model, which is often found to be important for continuous forecasts of economic activity (e.g., GDP growth), since we are interested in estimating business cycle turning points where modeling persistence in the data is likely to deteriorate the ability of the model to detect regime switches. In that respect, we follow for example Granger and Terasvirta (1999).

3 Combination schemes

In the empirical application, univariate and bivariate specifications each generate 50 estimates for the probability of recession (i.e., one for each U.S. state). This information is summarized using two different classes of combination schemes: Bayesian model averaging and dynamic model averaging.

3.1 Bayesian model averaging

3.1.1 Likelihood approach

Suppose that we have K different models, M_k for $k = 1, ..., K$, which all seek to explain y_t. The model M_k depends upon the regression parameters of the econometric specification (univariate or multivariate), collected in the vector Θ_k. Hence, the posterior distribution for the parameters calculated from model M_k can be written as:

$$f(\Theta_k|y_t, M_k) = \frac{f(y_t|\Theta_k, M_k)f(\Theta_k|M_k)}{f(y_t|M_k)}. \tag{6}$$

Analogously, as suggested by Koop (2003), if one is interested in comparing different models, we can use Bayes' rule to derive a probability statement about what we do not know (i.e., whether model M_k is appropriate or not to explain y_t) conditional on what we do know (i.e., the data, y_t). This implies that the posterior model probability can be used to assess the degree of support for model k:

$$f(M_k|y_t) = \frac{f(y_t|M_k)f(M_k)}{f(y_t)}, \tag{7}$$

where $f(y_t) = \sum_{j=1}^{K} f(y_t|M_j)f(M_j)$, $f(M_k)$ is the prior probability that model k is true and $f(y_t|M_k)$ is the marginal likelihood for model k. Following Newton and Raftery (1994), the marginal likelihood is calculated from the harmonic mean estimator, which is a simulation-consistent estimate that uses samples from the posterior density.[1] The harmonic mean estimator of the marginal likelihood is:

$$f(y_t|M_k) = \left(\frac{1}{N} \sum_{n=1}^{N} \frac{1}{f(y_t|M_k^{(n)})} \right)^{-1} \tag{8}$$

where $f(y_t|M_k^{(n)})$ is the posterior density available from simulation n, and N is the total number of simulations. Initially, one could assume that all models are equally likely, that

[1] Note that alternative approaches could be used to calculate the marginal likelihood (see, e.g., Chib (1995) or Fruhwirth-Schnatter (2004)). However, these alternative methods are typically computationally demanding in that they require a substantial increase in the number of simulations, which is not suitable in our empirical application, since we have to estimate many models in a recursive out-of-sample forecasting experiment.

is $f(M_k) = \frac{1}{K}$. Alternatively, one could use the employment share of each U.S. state to set the prior probability for each model. In the case of equal prior probability for each model, the weights for model k are simply given as:

$$f(M_k|y_t) = \frac{f(y_t|M_k)}{\sum_{j=1}^{K} f(y_t|M_j)} \tag{9}$$

3.1.2 Combined approach

Given that our models are designed to predict NBER recessions rather than predicting the national activity indicator y_t, an alternative weighting scheme could be implemented to reflect this objective. Indeed, we can rely on Bayes' rule to derive a probability statement about the most appropriate model M_k to explain the regimes S_t conditional on the data and the estimated probability of being in a given regime derived from the Hamilton filter for Markov-switching models, $P(S_t|y_t)$.[2] Therefore, the posterior model probability can be expressed as:

$$f(M_k|y_t, S_t) = \frac{f(y_t, S_t|M_k)f(M_k)}{f(y_t, S_t)} \tag{10}$$

$$= \frac{f(S_t|y_t, M_k)f(y_t|M_k)f(M_k)}{f(S_t|y_t)f(y_t)} \tag{11}$$

where $f(S_t|y_t)P(y_t) = \sum_{j=1}^{K} f(S_t|y_t, M_j)f(y_t|M_j)f(M_j)$, $f(M_k)$ is the prior probability that model k is true, $f(y_t|M_k)$ is the marginal likelihood for model k, and $f(S_t|y_t, M_k)$ indicates the model's ability to fit the business cycle regimes. We use the inverse Quadratic Probability Score (QPS) to evaluate $f(S_t|y_t, M_k)$, since the QPS is the most common measure to evaluate discrete outcomes in the business cycle literature.[3] The QPS associated to model k is defined as follows:

$$QPS_k = \frac{2}{T}\sum_{t=1}^{T}(P(S_t^k = 0|\psi_t) - NBER_t)^2, \tag{12}$$

where $P(S_t^k = 0|\psi_t)$ is the probability of being in recession, given information up to t, ψ_t, and $NBER_t$ is a dummy variable that takes on a value of 1 if the U.S. economy is in recession at time t according to the NBER business cycle dating committee and 0 otherwise. QPS is bounded between 0 and 2, and perfect predictions yield a QPS of 0. Hence, the lower the QPS, the better the ability of the model to fit the U.S. business cycle is. Accordingly, the posterior model probability for model k reads as:

$$f(M_k|y_t, S_t) = \frac{f(y_t|M_k)f(M_k)QPS_k^{-1}}{\sum_{j=1}^{K} f(y_t|M_j)f(M_j)QPS_j^{-1}}. \tag{13}$$

[2]For ease of exposition, here, we only present the case of one single Markov chain driving the changes in the parameters of the model. The derivations can be relatively easily extended to the case of multiple Markov chains, but this would come at the cost of a much more demanding notation.

[3]Note that alternative criteria could be used to evaluate the models' ability to classify regimes. For example, the logarithmic probability score or the area under the Receiver Operating Characteristics (see, e.g., Berge and Jordà (2011)) could be used. However, to streamline the paper we exclusively use the QPS, which is the most commonly used criteria to evaluate discrete outcomes.

One could use the U.S. employment share of each state as prior probability for each model or equal prior weights. In the case of equal prior probability for each model, the posterior probability is:

$$f(M_k|y_t, S_t) = \frac{\eta_k}{\sum_{j=1}^{K} \eta_j}, \tag{14}$$

where

$$\eta_k = \frac{f(y_t|M_k)}{QPS_k}. \tag{15}$$

3.1.3 QPS approach

Notice that the posterior model probability in Equation (10) focuses on a joint fit of data, y_t, and business cycle phases, S_t. However, since we are only interested in assessing the ability of model M_k in explaining the business cycle phases, S_t, we avoid conditioning on y_t and following the reasoning in Section 3.1.2, propose the following expression for the posterior probability model:

$$f(M_k|S_t) = \frac{f(S_t|M_k)f(M_k)}{\sum_{j=1}^{K} f(S_t|M_j)f(M_j)} \tag{16}$$

$$= \frac{f(M_k)QPS_k^{-1}}{\sum_{j=1}^{K} f(M_j)QPS_j^{-1}}. \tag{17}$$

In the case of equal prior probability for each model, the posterior probability is given by the normalized inverse QPS:

$$f(M_k|y_t) = \frac{QPS_k^{-1}}{\sum_{j=1}^{K} QPS_j^{-1}}. \tag{18}$$

3.2 Dynamic model averaging

3.2.1 Raftery's approach

Dynamic model averaging originates from the work of Raftery et al. (2010), and has been first implemented in econometrics by Koop and Korobilis (2012) and Koop and Korobilis (2013).

To compute weights that vary over time associated to model k, we only need the predictive density at time t of the corresponding model, $f_k(y_t|\psi_{t-1})$, and a coefficient referred to as the forgetting factor, α. Denote $\pi_{t-1|t-1,k}$ the predicted probability that model k is the most appropriate for forecasting at time $t-1$ given information up to time $t-1$. Raftery et al. (2010) suggest that predictions of $\pi_{t-1|t-1,k}$ can be done as follows:

$$\pi_{t|t-1,k} = \frac{\pi_{t-1|t-1,k}^{\alpha}}{\sum_{j=1}^{K} \pi_{t-1|t-1,j}^{\alpha}} \tag{19}$$

where $0 < \alpha < 1$ is set to a fixed value slightly less than one. The forgetting factor α is the coefficient that governs the amount of persistence in the models' weights. The higher the α, the higher the weight attached to past predictive performance is. Following Raftery et al. (2010), we first set $\alpha = 0.99$, which implies that forecasting performance from two years ago receives about 78.5 per cent weight compared with last period's forecasting performance. We also report results with $\alpha = 0.95$ so as to give lower weights to past forecasting performance (in this case, information from two years ago receives about 29 per cent weight compared with last period's information).

Once y_t is observed, $\pi_{t|t-1,k}$, can be updated by using the predictive density, as follows:

$$\pi_{t|t,k} = \frac{\pi_{t|t-1,k} f_k(y_t|\psi_{t-1})}{\sum_{j=1}^{K} \pi_{t|t-1,j} f_j(y_t|\psi_{t-1})} \tag{20}$$

The last two equations are repeated consecutively for each t, starting with equal weight for each model at $t = 1$.

Dynamic model averaging differs from Bayesian model averaging in that no simulation is required to calculate the models' weights, and the weights vary over time ($t = \{1, ..., T\}$). A detailed explanation about the algorithm used to calculate DMA weights in the context of Markov-switching models is presented in Appendix B.

3.2.2 Combined approach

In line with Section 3.1.2, we also allow for the possibility that both, the marginal likelihood and the cumulative QPS, could inform about the model's ability to predict business cycle phases. Therefore, the updating equation reads as:

$$\pi_{t|t,k} = \frac{\pi_{t|t-1,k} \eta_{t|t-1,k}}{\sum_{j=1}^{K} \pi_{t|t-1,j} \eta_{t|t-1,j}}, \tag{21}$$

where

$$\eta_{t|t-1,k} = \frac{f_k(y_t|\psi_{t-1})}{Q_{t|t,k}}, \tag{22}$$

and $Q_{t|t,k}$ is the cumulative QPS at time t for model k defined as:

$$Q_{t|t,k} = \frac{2}{t}(\sum_{\tau=1}^{t} P(S_\tau^k = 0|\psi_\tau) - NBER_\tau)^2. \tag{23}$$

The model prediction equation remains the same as in equation (19).

3.2.3 QPS approach

Again, since we are only interested in predicting business cycle phases instead of forecasting the national activity variable, we modify Raftery et al. (2010) approach. Specifically, in line with Section 3.1.3, in the updating equation, we replace the marginal likelihood, which measures how well the model fits the data, with a measure of goodness-of-fit for business cycle regimes. Hence, the updating equation reads as:

$$\pi_{t|t,k} = \frac{\pi_{t|t-1,k} Q_{t|t,k}^{-1}}{\sum_{j=1}^{K} \pi_{t|t-1,j} Q_{t|t,j}^{-1}}. \tag{24}$$

4 Simulation Study

We conduct a Monte Carlo experiment to study in a controlled set-up the validity of the different model averaging schemes detailed in the previous section. In doing so, we choose a data generating process that closely mimics the empirical application of the paper. Equations (25) to (27) detail the data generating process. First, the dependent variable y_t is generated according to the following equation:

$$y_t = \mu_0^y + \mu_1^y S_t + \varepsilon_t^y, \tag{25}$$

where $\varepsilon_t^y \sim N(0, \sigma_y^2)$, and $(\mu_0^y, \mu_1^y) = (-1, 2)$.

The $x_{k,t}$'s variables are instead generated from the following equation:

$$x_{k,t} = \mu_0^k + \mu_1^k S_t + \sigma_k \varepsilon_t^k, \text{ for } k = \{1, ..., K\}. \tag{26}$$

where $\varepsilon_t^k \sim N(0, \sigma_k^2)$.[4]

The intercepts for the $x_{k,t}$'s variables are given by:

$$\mu_j^k = \mu_j^y + \mu_j^y \epsilon_{k,j}, \text{ for } j = \{0, 1\}. \tag{27}$$

where $\epsilon_{k,j} \sim U(-1, 1)$, so that the intercepts for the x_t's variables are closely related to the intercepts of the variable y_t.

While the intercepts' values μ_0^y and μ_1^y are kept constant, we use four different values for the variance of the innovations ($\sigma^2 = \{0.5, 1, 1.5, 2\}$). Moreover, S_t is a standard first-order Markov chain with two regimes and constant transition probabilities given by: (p_{00}, p_{11})=$(0.8, 0.9)$. In this way, the first regime is associated with a negative growth rate and it is less persistent than the second regime, a common feature of business cycle series that typically exhibit different regimes' duration. Finally, the series are generated with length $T = 200$ and the number of x_t variables is set to $K = 20$. The total number of replications is set to 1000. For each replication, the total number of simulations to estimate the model's parameters is 3000, discarding the first 1000 simulations to account for start-up effects.

Table 1 reports, for the different model-averaging schemes and under the different scenarios considered (i.e., BMA, DMA and an equal-weight scheme), the average in-sample QPS obtained across the 1000 replications.[5] For ease of computations, we also assume that a single Markov chain S_t drives the changes for both the y_t and x_t's variables.[6]

The results show that, first, in both univariate and bivariate cases, the lowest QPS' are obtained when using the QPS-based model averaging scheme. This holds true for both

[4] Note that the variance of the innovations is the same for both the y_t and the x_t's (i.e., $\sigma_x = \sigma_y$) so as to avoid large differences in volatility across series.

[5] We only consider in-sample Monte-Carlo experiments owing to the too demanding computational task that would be required for a fully recursive out-of-sample exercise.

[6] This is not too detrimental since our primary objective is to estimate turning points at the national or aggregate level, hence we do not loose much in assuming a single Markov-chain driving the parameter changes in the bivariate case.

Table 1: Monte Carlo simulation results

	σ	0.5	1	1.5	2
Panel A: Univariate model					
Dynamic Model	Likelihood-based	0.447	0.360	0.351	0.358
Averaging	QPS-based	**0.027**	**0.161**	**0.257**	**0.311**
	Combined	0.027	0.164	0.261	0.314
Bayesian Model	Likelihood-based	0.131	0.347	0.370	0.376
Averaging	QPS-based	0.055	0.223	0.297	0.335
	Combined	0.114	0.340	0.368	0.374
Equal weight		0.443	0.433	0.458	0.476
Panel B: Bivariate model					
Dynamic Model	Likelihood-based	0.034	0.164	0.268	0.325
Averaging	QPS-based	**0.016**	**0.137**	**0.248**	**0.310**
	Combined	0.016	0.144	0.252	0.314
Bayesian Model	Likelihood-based	0.016	0.147	0.266	0.324
Averaging	QPS-based	0.016	0.160	0.273	0.329
	Combined	0.016	0.147	0.266	0.324
Equal weight		0.271	0.366	0.442	0.472

Note: This table reports the QPS averaged over 1000 replications using the different combination schemes outlined in Section 3. Bold entries in each panel indicate the lowest QPS for a selected DGP. See text for full details about the design of the Monte Carlo experiment.

DMA and BMA in the univariate case, and the differences are the most noticeable in the BMA context. Second, in the context of DMA, the combined weighting scheme that relies on both the QPS and the marginal likelihood is a very close second best weighting scheme, which further emphasizes the value of the QPS to calculate models' weights. Third, as the volatility of the series increases (i.e., for higher values of σ), the differences in terms of QPS across the weighting schemes tend to soften. This is relatively intuitive in that, given the DGP's we consider, as the volatility of the series increases, regimes shifts in the series become less apparent, and it is therefore more difficult to make inference on the regimes, which translates into higher QPS, and lower value added resulting from weighting schemes based on QPS. Overall, this simulation exercise underlines the relevance of our model-averaged scheme based on past predictive performance to classify the regimes (i.e., QPS-based). The next section evaluates the relevance of this framework from an empirical point of view, forecasting national U.S. recessions based on a set of regional indicators.

5 Empirical Results

5.1 Data

We use alternatively industrial production and employment data as a measure of national economic activity. These two indicators are available on a monthly basis, and are frequently considered as important measures of economic activity in the U.S. The state-level data we use are the employees on non-farm payrolls data series published at a monthly frequency for each U.S. state by the U.S. Bureau of Labor Statistics. These data are available on a not seasonally adjusted basis since at least January 1960 for all U.S. states. In contrast, data on a seasonally adjusted basis are available since January 1990, and real-time data vintages are only available since June 2007 from the "Alfred" real-time database of the Federal Reserve Bank of St. Louis.[7] All data are taken as 100 times the change in the log-level of the series to obtain monthly percent changes. To facilitate inference on the regimes, and obtain a long enough evaluation sample to assess the accuracy of the forecasts, we use data starting from 1960, and the data are appropriately seasonally adjusted. Hence, The full estimation sample extends from February 1960 to April 2014.

5.2 In-sample results

The in-sample results are based on the data vintage from May 2014 with last observation for April 2014. For brevity, we only report the results for the models with national employment data as a dependent variable.[8] All models are estimated discarding the first 2000 replications to account for start-up effects, running 5000 additional simulations to calculate the posterior distribution of parameters (see Appendix A for additional details). To assess the ability of regime-switching models to predict U.S. recessions, we use the in-sample Quadratic Probability Score (QPS_k^{IS}) defined as:

$$QPS_k^{IS} = \frac{2}{T}\sum_{t=1}^{T}(P(S_t^k = 0|\psi_t) - NBER_t)^2 \tag{28}$$

where T is the size of the full sample, $P(S_t^k = 0|\psi_t)$ is the probability of being in a low mean regime (i.e., the recession regime), and $NBER_t$ is a dummy variable that takes on a value of 1 if the U.S. economy is in recession according to the NBER business cycle dating committee and 0 otherwise.

Table 2 reports the in-sample parameter estimates for all individual models in the univariate case, as well as their quadratic probability scores. First, all univariate models exhibit a classical cycle for employment in that average growth in the low mean regime

[7]Data are available on `http://alfred.stlouisfed.org/`, and are typically available with a roughly three-week delay for the state-level data, about a 1-week delay for national employment, and a 2-week delay for industrial production.

[8]The main conclusions are relatively unchanged when using industrial production as a dependent variable. Detailed results are available upon request.

(i.e., μ_0) is always negative, whereas average growth in the high mean regime (i.e., $\mu_0 + \mu_1$) is always positive. There are also little differences for the intercept estimates across all models. However, differences are noticeable for the slope parameter β. For example, perhaps unsurprisingly, the lowest slope parameter is for the model using employment data for Alaska. In contrast, the highest slope parameter is for the model with employment data for the state of Ohio. In addition, models with employment data for the states of New York, Pennsylvania, New Jersey or California also yield large slope parameters, suggesting the importance of the employment data from these states to explain the national U.S. employment data. Finally, the model with employment data for the state of Virginia yields the lowest (in-sample) QPS, whereas the model with the highest QPS is the one using employment data for the state of Ohio. This suggests that the most relevant model for explaining aggregate U.S. employment growth is not necessarily the most relevant for estimating U.S. business cycle regimes.

Table 3 and Table 4 report the results for the bivariate models. First, as expected, the intercepts for the equation on U.S. employment vary little across models and are roughly in line with the parameter estimates from the univariate models. Second, for four states (Alaska, Arizona, North Dakota, and New Mexico), the intercepts for the state employment growth are positive in both regimes, that is the bivariate model estimates growth cycle rather than classical cycle for the dynamics of state employment. Third, the lowest in-sample QPS is obtained from the model using New Jersey employment data, followed by the model with Maryland employment data.

Table 5 reports the in-sample QPS with the different combination schemes outlined in Section 3 using alternatively employment and industrial production data as a measure of national economic activity. First, models with industrial production yield lower QPS compared with models with employment data. Second, in the univariate case, the best specification is obtained by the MS-AR model with industrial production followed by the model with industrial production and weights obtained from DMA using a combination of predictive likelihood and QPS. Third, for multivariate models, models with industrial production also tend to yield lower QPS. In detail, the equal weight specification produced the lowest QPS followed by the DMA combination scheme based on the QPS. Fourth, for DMA combination schemes, a lower value for the forgetting factor α tends to yield lower QPS. Figure 1 reports the probability of recession from selected models, which shows that these models can track very well the recessions defined by the NBER business cycle dating committee. One can also see that models using employment data as a measure of national economic activity identified the last three recessions as being longer than the NBER recession estimates. This is not surprising given that these recessions were associated with a jobless recovery.

To better understand the results from Table 5, Figures 2, 3 and 4 show the weights attached to each individual model with the dynamic model averaging (DMA) scheme in the univariate case. Figure 2 reports the results from the standard dynamic model averaging scheme where the weights are based exclusively on the predictive likelihood. In the case of employment as a dependent variable (Panel A of Figure 1), Ohio gets a probability of inclusion close to one for nearly the entire sample, except in the 1990's where the states of New Jersey and New York also exhibit a non-negligible probability of inclusion (and also Florida at the end of the sample). In the case of industrial production, the weights given to individual models are more even across the different models, except at the end of the sample where the states of Virginia and Florida get a predominant weight. Figure 3 (i.e., where the DMA weights are based exclusively on past QPS) and Figure 4 (i.e., where the weights are based on a combination of past QPS and predictive likelihood) show a substantial time variation in the weights attached to individual models. In both Figure 3 and Figure 4, the weight attached to the model using Maryland employment data is high in the early part of the sample, whereas it is the model using data for the state of Virginia that gets the highest weight at the end of the sample (or the state of Idaho when using industrial production as a dependent variable, see panel B). Figure 5 reports the weights obtained from Bayesian model averaging (BMA) schemes in the univariate case. Panel A of Figure 5 shows that the model with Ohio employment data gets a weight of one with standard Bayesian model averaging, which is not surprising given that Table 1 showed that the model with Ohio employment data exhibited the highest correlation with the national employment data.[9] When explaining national industrial production, it is the employment data from the state of Michigan that gets a weight near 1 (see panel B). In contrast, BMA weights based on QPS yield larger weights to heavily populated states (e.g., California or New York).

[9] The fact that BMA tends to give a weight of 1 to a single model is not very surprising. Geweke and Amisano (2011) suggest to use the historical log predictive score to mitigate this issue. We also implemented this approach, but obtained results relatively close to BMA in that a single model obtained the largest weight with only few other models obtaining a non-negligible weight.

Table 2: In-sample parameter estimates - Univariate models

State	μ_0	μ_1	β	QPS		μ_0	μ_1	β	QPS
Alabama	-0.099	0.266	0.263	0.191	Montana	-0.149	0.352	0.082	0.179
	[-0.119,-0.079]	[0.244,0.287]	[0.241,0.285]			[-0.170,-0.128]	[0.330,0.374]	[0.065,0.098]	
Alaska	-0.152	0.370	0.000	0.169	Nebraska	-0.145	0.324	0.209	0.167
	[-0.175,-0.128]	[0.346,0.393]	[-0.011,0.010]			[-0.168,-0.122]	[0.300,0.347]	[0.183,0.235]	
Arizona	-0.140	0.282	0.181	0.169	Nevada	-0.152	0.325	0.090	0.160
	[-0.162,-0.117]	[0.258,0.307]	[0.174,0.201]			[-0.175,-0.129]	[0.301,0.350]	[0.073,0.106]	
Arkansas	-0.138	0.310	0.193	0.190	New Hampshire	-0.137	0.300	0.215	0.172
	[-0.158,-0.118]	[0.290,0.332]	[0.174,0.211]			[-0.169,-0.112]	[0.275,0.328]	[0.193,0.238]	
California	-0.111	0.240	0.349	0.145	New Jersey	-0.145	0.300	0.364	0.127
	[-0.135,-0.087]	[0.213,0.266]	[0.317,0.380]			[-0.170,-0.124]	[0.279,0.324]	[0.339,0.388]	
Colorado	-0.177	0.324	0.223	0.124	New Mexico	-0.141	0.322	0.166	0.179
	[-0.202,-0.150]	[0.299,0.349]	[0.199,0.248]			[-0.164,-0.118]	[0.297,0.345]	[0.143,0.191]	
Connecticut	-0.124	0.306	0.222	0.126	New York	-0.098	0.270	0.418	0.202
	[-0.157,-0.095]	[0.276,0.335]	[0.198,0.245]			[-0.131,-0.071]	[0.246,0.300]	[0.383,0.452]	
Delaware	-0.137	0.339	0.073	0.171	North Carolina	-0.115	0.243	0.325	0.154
	[-0.161,-0.113]	[0.315,0.363]	[0.061,0.085]			[-0.139,-0.090]	[0.218,0.267]	[0.300,0.349]	
Florida	-0.124	0.243	0.290	0.215	North Dakota	-0.149	0.353	0.071	0.176
	[-0.146,-0.103]	[0.222,0.265]	[0.267,0.313]			[-0.171,-0.126]	[0.330,0.376]	[0.049,0.093]	
Georgia	-0.112	0.234	0.290	0.152	Ohio	-0.023	0.168	0.452	0.247
	[-0.143,-0.087]	[0.210,0.262]	[0.295,0.341]			[-0.042,-0.004]	[0.150,0.187]	[0.430,0.472]	
Hawaii	-0.150	0.352	0.078	0.165	Oklahoma	-0.134	0.324	0.144	0.166
	[-0.173,-0.127]	[0.329,0.375]	[0.060,0.095]			[-0.162,-0.109]	[0.299,0.350]	[0.122,0.168]	
Idaho	-0.145	0.335	0.100	0.179	Oregon	-0.119	0.293	0.159	0.177
	[-0.166,-0.123]	[0.313,0.357]	[0.083,0.118]			[-0.142,-0.095]	[0.268,0.318]	[0.136,0.182]	
Illinois	-0.064	0.238	0.335	0.200	Pennsylvania	-0.072	0.246	0.384	0.218
	[-0.087,-0.043]	[0.217,0.261]	[0.309,0.360]			[-0.092,-0.054]	[0.227,0.267]	[0.357,0.411]	
Indiana	-0.073	0.237	0.287	0.220	Rhode Island	-0.113	0.307	0.172	0.177
	[-0.092,-0.055]	[0.217,0.257]	[0.269,0.306]			[-0.137,-0.090]	[0.284,0.330]	[0.153,0.191]	
Iowa	-0.122	0.301	0.222	0.175	South Carolina	-0.109	0.266	0.230	0.177
	[-0.142,-0.102]	[0.280,0.322]	[0.199,0.245]			[-0.131,-0.087]	[0.242,0.288]	[0.209,0.252]	
Kansas	-0.130	0.325	0.123	0.172	South Dakota	-0.150	0.344	0.123	0.179
	[-0.152,-0.108]	[0.303,0.347]	[0.105,0.141]			[-0.171,-0.128]	[0.323,0.366]	[0.102,0.144]	
Kentucky	-0.129	0.306	0.184	0.166	Tennessee	-0.109	0.261	0.274	0.198
	[-0.147,-0.109]	[0.286,0.326]	[0.168,0.200]			[-0.131,-0.089]	[0.241,0.283]	[0.255,0.293]	
Louisiana	-0.138	0.342	0.087	0.180	Texas	-0.124	0.257	0.315	0.187
	[-0.161,-0.115]	[0.319,0.365]	[0.069,0.104]			[-0.152,-0.098]	[0.232,0.284]	[0.284,0.345]	
Maine	-0.129	0.317	0.164	0.165	Utah	-0.143	0.316	0.148	0.167
	[-0.155,-0.104]	[0.292,0.342]	[0.141,0.186]			[-0.167,-0.120]	[0.291,0.340]	[0.125,0.171]	
Maryland	-0.159	0.329	0.206	0.131	Vermont	-0.135	0.323	0.134	0.172
	[-0.181,-0.136]	[0.305,0.351]	[0.184,0.227]			[-0.158,-0.112]	[0.299,0.346]	[0.115,0.154]	
Massachusetts	-0.134	0.305	0.268	0.121	Virginia	-0.171	0.301	0.309	0.097
	[-0.189,-0.087]	[0.267,0.351]	[0.234,0.298]			[-0.207,-0.137]	[0.272,0.330]	[0.282,0.337]	
Michigan	-0.103	0.300	0.131	0.201	Washington	-0.128	0.295	0.190	0.176
	[-0.121,-0.083]	[0.279,0.319]	[0.120,0.142]			[-0.150,-0.108]	[0.272,0.317]	[0.166,0.215]	
Minnesota	-0.107	0.256	0.305	0.179	Wisconsin	-0.112	0.278	0.258	0.183
	[-0.126,-0.088]	[0.236,0.277]	[0.280,0.329]			[-0.133,-0.090]	[0.255,0.301]	[0.232,0.282]	
Mississippi	-0.121	0.288	0.223	0.185	West Virginia	-0.142	0.358	0.035	0.176
	[-0.142,-0.100]	[0.266,0.310]	[0.204,0.242]			[-0.165,-0.119]	[0.335,0.381]	[0.027,0.043]	
Missouri	-0.120	0.299	0.230	0.187	Wyoming	-0.145	0.355	0.048	0.176
	[-0.138,-0.101]	[0.280,0.319]	[0.209,0.251]			[-0.168,-0.122]	[0.332,0.319]	[0.034,0.061]	

Note: μ_0 is the mean growth rate in recession for aggregate U.S. employment, $\mu_0 + \mu_1$ is the mean growth rate in expansions for aggregate U.S. employment, β is the parameter entering before the state-level employment data in equation (1). The parameter estimates are reported as the median over 5000 replications. The estimation sample extends from February 1960 to April 2014. QPS is the Quadratic Probability Score for individual models as defined in equation (28), and the 90 per cent coverage intervals are reported in brackets.

Table 3: In-sample Parameter estimates - Multivariate models

State	μ_0	μ_1	QPS		μ_0	μ_1	QPS
Alabama	-0.097	0.306	0.192	Montana	-0.134	0.349	0.178
	[-0.128,-0.059]	[0.262,0.340]			[-0.155,-0.112]	[0.327,0.371]	
	-0.128	0.317			-1.870	2.040	
	[-0.911,-0.042]	[0.228,1.055]			[-2.234,-1.541]	[1.711,2.403]	
Alaska	-0.150	0.368	0.169	Nebraska	-0.114	0.326	0.161
	[-0.173,-0.125]	[0.344,0.392]			[-0.143,-0.089]	[0.301,0.354]	
	0.161	2.115			-0.838	0.995	
	[0.125,0.197]	[1.933,2.302]			[-1.178,-0.017]	[0.209,1.330]	
Arizona	-0.101	0.308	0.189	Nevada	-0.134	0.349	0.178
	[-0.126,-0.077]	[0.284,0.334]			[-0.158,-0.111]	[0.325,0.373]	
	0.133	0.403			-0.125	0.640	
	[0.085,0.176]	[0.356,0.449]			[-0.261,-0.016]	[0.543,0.756]	
Arkansas	-0.104	0.310	0.186	New Hampshire	-0.131	0.340	0.182
	[-0.126,-0.082]	[0.288,0.332]			[-0.157,-0.103]	[0.311,0.369]	
	-1.641	1.829			-0.227	0.479	
	[-1.975,-1.214]	[1.394,2.164]			[-0.330,-0.142]	[0.400,0.568]	
California	-0.118	0.326	0.181	New Jersey	-0.096	0.289	0.129
	[-0.144,-0.090]	[0.299,0.353]			[-0.124,-0.070]	[0.264,0.319]	
	-0.099	0.362			-0.795	0.908	
	[-0.133,-0.066]	[0.326,0.396]			[-0.966,-0.296]	[0.434,1.076]	
Colorado	-0.121	0.332	0.179	New Mexico	-0.122	0.334	0.168
	[-0.145,-0.097]	[0.309,0.356]			[-0.146,-0.096]	[0.310,0.359]	
	-0.056	0.374			0.096	0.232	
	[-0.100,-0.015]	[0.333,0.418]			[0.018,0.140]	[0.186,0.280]	
Connecticut	-0.125	0.337	0.169	New York	-0.112	0.314	0.178
	[-0.153,-0.097]	[0.309,0.365]			[-0.147,-0.077]	[0.281,0.347]	
	-0.184	0.341			-0.113	0.220	
	[-0.239,-0.130]	[0.286,0.397]			[-0.163,-0.073]	[0.180,0.268]	
Delaware	-0.127	0.341	0.167	North Carolina	-0.136	0.341	0.185
	[-0.153,-0.102]	[0.317,0.366]			[-0.161,-0.110]	[0.313,0.366]	
	-1.910	2.088			-0.231	0.484	
	[-2.405,-1.485]	[1.665,2.581]			[-0.280,-0.179]	[0.430,0.535]	
Florida	-0.087	0.304	0.238	North Dakota	-0.133	0.352	0.182
	[-0.119,-0.060]	[0.277,0.334]			[-0.157,-0.110]	[0.329,0.376]	
	-0.064	0.436			0.121	0.563	
	[-0.111,-0.020]	[0.390,0.483]			[0.100,0.142]	[0.481,0.640]	
Georgia	-0.117	0.322	0.203	Ohio	0.013	0.166	0.252
	[-0.142,-0.091]	[0.295,0.348]			[-0.008,0.034]	[0.147,0.185]	
	-0.172	0.448			-1.473	1.552	
	[-0.226,-0.119]	[0.393,0.504]			[-1.734,-1.203]	[1.284,1.814]	
Hawaii	-0.135	0.351	0.163	Oklahoma	-0.133	0.348	0.167
	[-0.159,-0.110]	[0.327,0.375]			[-0.160,-0.105]	[0.321,0.373]	
	-2.015	2.209			-0.218	0.453	
	[-2.521,-1.575]	[1.770,2.709]			[-0.273,-0.161]	[0.397,0.510]	
Idaho	-0.143	0.359	0.171	Oregon	-0.155	0.368	0.162
	[-0.165,-0.120]	[0.335,0.382]			[-0.181,-0.130]	[0.343,0.394]	
	-0.316	0.598			-0.316	0.598	
	[-0.465,-0.165]	[0.481,0.753]			[-0.386,-0.255]	[0.536,0.666]	

Note: This table reports results from the estimation of equation (5). μ_0 is the mean growth rate in recession, $\mu_0 + \mu_1$ is the mean growth rate in expansions. For each state, the first row indicates the results for employment at the national level, whereas the second row indicates results for employment at the state-level. The parameter estimates are reported as the median over 5000 replications. The estimation sample extends from February 1960 to April 2014. QPS is the Quadratic Probability Score for individual models as defined in equation (28), and 90 per cent coverage intervals are reported in brackets.

Table 4: In-sample Parameter estimates - Multivariate models (cont'd)

State	μ_0	μ_1	QPS		μ_0	μ_1	QPS
Illinois	-0.106	0.317	0.196	Pennsylvania	-0.095	0.301	0.191
	[-0.132,-0.081]	[0.291,0.344]			[-0.125,-0.066]	[0.269,0.332]	
	-0.156	0.304			-0.156	0.263	
	[-0.196,-0.116]	[0.263,0.346]			[-0.237,-0.106]	[0.212,0.328]	
Indiana	-0.050	0.249	0.206	Rhode Island	-0.108	0.315	0.175
	[-0.078,-0.027]	[0.226,0.278]			[-0.131,-0.084]	[0.291,0.339]	
	-0.508	0.640			-1.101	1.199	
	[-0.755,-0.342]	[0.485,0.880]			[-1.509,-0.836]	[0.941,1.601]	
Iowa	-0.086	0.289	0.166	South Carolina	-0.136	0.348	0.183
	[-0.108,-0.064]	[0.268,0.311]			[-0.159,-0.113]	[0.323,0.371]	
	-1.353	1.494			-0.237	0.499	
	[-1.539,-1.184]	[1.322,1.679]			[-0.318,-0.175]	[0.434,0.576]	
Kansas	-0.108	0.318	0.166	South Dakota	-0.135	0.352	0.178
	[-0.132,-0.084]	[0.295,0.342]			[-0.159,-0.112]	[0.328,0.375]	
	-3.517	3.661			-0.017	0.248	
	[-3.976,-3.045]	[3.191,4.119]			[-0.125,0.066]	[0.175,0.341]	
Kentucky	-0.088	0.292	0.162	Tennessee	-0.108	0.317	0.166
	[-0.108,-0.067]	[0.271,0.313]			[-0.133,-0.084]	[0.291,0.344]	
	-1.178	1.359			-0.105	0.334	
	[-1.438,-0.969]	[1.153,1.615]			[-0.164,-0.048]	[0.274,0.394]	
Louisiana	-0.117	0.332	0.180	Texas	-0.112	0.329	0.201
	[-0.142,-0.093]	[0.309,0.356]			[-0.138,-0.086]	[0.304,0.354]	
	-4.004	4.157			-0.050	0.365	
	[-4.342,-3.664]	[3.816,4.495]			[-0.086,-0.014]	[0.330,0.401]	
Maine	-0.108	0.316	0.163	Utah	-0.139	0.353	0.167
	[-0.135,-0.082]	[0.291,0.342]			[-0.163,-0.114]	[0.328,0.377]	
	-2.373	2.498			-0.127	0.459	
	[-2.787,-1.934]	[2.057,2.913]			[-0.189,-0.068]	[0.404,0.517]	
Maryland	-0.126	0.330	0.137	Vermont	-0.111	0.320	0.174
	[-0.150,-0.102]	[0.307,0.356]			[-0.134,-0.086]	[0.296,0.385]	
	-0.959	1.135			-1.623	1.791	
	[-1.278,-0.191]	[0.387,1.452]			[-2.087,-1.178]	[1.344,2.253]	
Massachusetts	-0.107	0.319	0.194	Virginia	-0.119	0.330	0.184
	[-0.135,-0.080]	[0.293,0.347]			[-0.144,-0.094]	[0.303,0.356]	
	-0.264	0.410			-0.039	0.295	
	[-0.317,-0.213]	[0.359,0.463]			[-0.098,0.017]	[0.240,0.350]	
Michigan	-0.086	0.292	0.197	Washington	-0.132	0.345	0.167
	[-0.107,-0.065]	[0.270,0.313]			[-0.157,-0.108]	[0.321,0.369]	
	-3.291	3.403			-0.112	0.412	
	[-3.666,-2.895]	[3.007,3.773]			[-0.162,-0.064]	[0.364,0.461]	
Minnesota	-0.114	0.324	0.186	Wisconsin	-0.133	0.345	0.186
	[-0.138,-0.088]	[0.300,0.350]			[-0.158,-0.109]	[0.320,0.371]	
	-0.096	0.325			-0.178	0.371	
	[-0.140,-0.053]	[0.280,0.371]			[-0.230,-0.127]	[0.318,0.425]	
Mississippi	-0.077	0.313	0.188	West Virginia	-0.133	0.351	0.177
	[-0.136,-0.077]	[0.279,0.345]			[-0.158,-0.109]	[0.320,0.371]	
	-0.068	0.304			-6.175	6.286	
	[-1.692,0.044]	[0.225,1.856]			[-6.708,-5.640]	[5.748,6.815]	
Missouri	-0.095	0.299	0.178	Wyoming	-0.143	0.360	0.171
	[-0.116,-0.074]	[0.277,0.321]			[-0.167,-0.117]	[0.336,0.385]	
	-1.038	1.160			-0.751	0.981	
	[-1.347,-0.866]	[0.991,1.466]			[-0.958,0.574]	[0.811,1.183]	

Note: See Table 3.

Table 5: In-sample Quadratic Probability Score

		QPS	
		Univariate model	Multivariate model
		Employment data	
Dynamic Model	Likelihood-based	0.226	0.226
Averaging	QPS-based	0.159	0.174
($\alpha = 0.99$)	Combined	0.136	0.173
Dynamic Model	Likelihood-based	0.154	0.187
Averaging	QPS-based	0.139	0.166
($\alpha = 0.95$)	Combined	0.122	0.182
Bayesian Model	Likelihood-based	0.236	0.178
Averaging	QPS-based	0.134	0.168
	Combined	0.236	0.178
MS-AR model		0.170	
Equal-weight		0.155	0.155
		Industrial Production	
Dynamic Model	Likelihood-based	0.102	0.182
Averaging	QPS-based	0.101	0.104
($\alpha = 0.99$)	Combined	0.099	0.126
Dynamic Model	Likelihood-based	0.120	0.172
Averaging	QPS-based	0.100	0.098
($\alpha = 0.95$)	Combined	0.095	0.119
Bayesian Model	Likelihood-based	0.121	0.194
Averaging	QPS-based	0.099	0.104
	Combined	0.121	0.194
MS-AR model		0.073	
Equal-weight		0.119	0.062

Note: This table reports the in-sample Quadratic Probability Score (QPS) for estimating U.S. business cycle turning points from univariate and multivariate models using different model-averaging schemes. α is the value of the forgetting factor when using dynamic model averaging schemes. The full estimation sample extends from February 1960 to April 2014. We discarded the first 2000 replications to account for start-up effects, and used the last 5000 replications to calculate all statistics.

5.3 Out-of-sample results

5.3.1 Full evaluation sample

The first estimation sample extends from February 1960 to December 1978, and it is recursively expanded until September 2013, that is the evaluation sample covers the period ranging from January 1979 to March 2014 (i.e., the last forecast six-month ahead refers to the month of March 2014). As such, our evaluation sample includes five recessions that covers 13.2 per cent of the sample. Such a long evaluation permits to mitigate the risks of spurious forecasting results. The models are re-estimated every month as new information becomes available.

We formulate forecasts for horizon $h = \{0, 1, 2, 3, 6\}$, that is from the current month ($h = 0$) up to six-month ahead ($h = 6$). We use the quadratic probability score (QPS) to evaluate the accuracy in predicting turning points. The out-of-sample QPS (QPS^{OOS}) is defined as follows:

$$QPS_k^{OOS} = \frac{2}{T - T_0 + 1} \sum_{t=T_0}^{T} (P(S_{t+h}^k = 0|\psi_t) - NBER_{t+h})^2 \tag{29}$$

where $T - T_0 + 1$ is the size of the evaluation sample, $P(S_{t+h}^k = 0|\psi_t)$ is the probability of being in the first regime (i.e., the recession regime) in period $t + h$, and $NBER_{t+h}$ is a dummy variable that takes on a value of 1 if the U.S. economy is in recession in period $t + h$ and 0 otherwise.

In comparing models, we also report results obtained from using the anxious index from the Survey of Professional Forecasters (SPF) of the Philadelphia Federal Reserve Bank. This index corresponds to the probability of a decline in real GDP. It is only available on a quarterly basis, but we disaggregate it at the monthly frequency assuming that its monthly value is constant over the three months of the quarter. Moreover, we also evaluate the statistical significance of our results using the Diebold-Mariano-West test to test for equal out-of-sample predictive accuracy (see Diebold and Mariano (1995) and West (1996)), using the likelihood-based weighting scheme as a benchmark model. In this way, we can evaluate from a statistical point of view the relevance of our weighting scheme based on the QPS compared with the traditional approach that relies exclusively on the likelihood.

Table 6 reports the results for the univariate models and Table 7 displays the results for the multivariate models. First, for univariate models, the combination scheme with industrial production using DMA weights based on the QPS obtains the best forecasting results for forecast horizons $h = \{0, 1, 2\}$, and the SPF anxious index obtained the best results for forecast horizons $h = \{3, 6\}$. Second, for multivariate models, the best results are obtained by the model using industrial production and DMA weights based on the QPS for forecast horizons $h = \{0, 1, 2\}$, and a combination of the predictive likelihood and QPS for forecast horizons $h = \{3, 6\}$. Third, the QPS-based combination schemes nearly always outperforms the combination schemes based on the likelihood only, and typically in a statistically significant way.

Figure 6 reports the one-month-ahead predicted probability of being in a recession from selected specifications. It shows that QPS-based DMA combination schemes perform well in that they capture very well all U.S. recessions. However, an important caveat of the out-of-sample analysis so far is that we only used revised data. In the next sub-section, we move to a real-time forecasting setting, concentrating on the prediction of the 2008-2009 recession.[10]

5.3.2 A closer look at the Great Recession

Revisions to macroeconomic data are substantial (see e.g. Croushore and Stark (2001)). Using data as available at the time the forecasts are made is therefore critical to evaluate realistically the models' forecasting ability. Real-time employment data are available for all 50 states starting from the June 2007 vintage with last observation for May 2007. Hence, our first estimation sample extends from February 1960 to May 2007, and it is recursively expanded until August 2013. As a result, the evaluation sample extends from May 2007 to August 2013, that is 76 months. In this case, since our evaluation sample only covers a limited period of time and only one recession, we do not calculate QPS statistics, but instead report the probability of being in a recession - defined as the last estimate available for the average probability of being in a recession (i.e., $P(\mathbf{S}_t = 0|\psi_t)$ where t is the last observation in the estimation sample) - and compare it with a number of alternatives.

Figure 7 reports the results for selected specifications using the QPS-based weighting scheme along with the probability of recession derived from the SPF Anxious index. In detail, this figure shows that the model using the employment data as a measure of national economic activity provides a timely update of the beginning of the recession in that the probability of recession is above 0.5 as early as April 2008. However, this model detects only with a substantial lag the end of the recession owing to the very slow recovery in labor market conditions. In contrast, the model using industrial production as a measure of national economic activity provides an accurate signal for the end of recession, but provides a late call for the beginning of the recession. Interestingly, the performance of the SPF anxious index is somewhat inferior to these two models despite the fact that the SPF uses a much larger information set than our model-based estimates. In particular, the anxious index provides a call of recession later than the model using national employment data and detects the end of the recession later than the model using national industrial production data. Overall, this suggests that employment data are very helpful to detect the beginning of recessions, whereas industrial production data rather provide valuable information about the end of recessions.

[10] As a robustness check, we also calculated QPS exclusively over the recession periods identified by the NBER (the results are not reported for space constraints). Over this restricted sample, the most accurate predictions at short-forecasting horizons are obtained by models using national employment and combining information with DMA based on the QPS. As such, this broadly confirms the full sample estimates in that weighting schemes based on the QPS provide valuable information.

Table 6: Out-of-sample Quadratic Probability Score - Univariate models

	Employment					
	Forecast horizon (months)	0	1	2	3	6
Dynamic Model	Likelihood-based	0.315	0.307	0.302	0.300	0.292
Averaging	QPS-based	0.184***	0.197***	0.214***	0.227***	0.252**
$\alpha = 0.99$	Combined	0.192***	0.207***	0.222***	0.233***	0.251**
Dynamic Model	Likelihood-based	0.222	0.236	0.246	0.252	0.259
Averaging	QPS-based	0.176	0.193*	0.210*	0.223	0.247
$\alpha = 0.95$	Combined	0.185***	0.206***	0.222**	0.231**	0.248
Bayesian Model	Likelihood-based	0.374	0.359	0.348	0.340	0.321
Averaging	QPS-based	0.196***	0.213***	0.229***	0.240***	0.256***
	Combined	0.374	0.359	0.348	0.340	0.321
Equal weight		0.209***	0.223***	0.237***	0.248***	0.263***
	Industrial Production					
Dynamic Model	Likelihood-based	0.239	0.240	0.243	0.248	0.252
Averaging	QPS-based	0.108***	0.136***	0.165**	0.188**	0.227
$\alpha = 0.99$	Combined	0.108***	0.135***	0.165**	0.187**	0.227
Dynamic Model	Likelihood-based	0.216	0.221	0.228	0.235	0.243
Averaging	QPS-based	**0.101*** **	**0.133*** **	**0.164** **	0.187*	0.227
$\alpha = 0.95$	Combined	0.110***	0.138***	0.166**	0.189**	0.227
Bayesian Model	Likelihood-based	0.209	0.219	0.230	0.236	0.247
Averaging	QPS-based	0.148***	0.171**	0.193**	0.209**	0.231
	Combined	0.208*	0.218*	0.230	0.236	0.247
Equal weight		0.131***	0.158***	0.182**	0.201**	0.229
SPF Anxious Index		0.141	0.161	0.180	**0.186**	**0.226**
MS-AR (Employment)		0.210	0.222	0.237	0.249	0.268
MS-AR (IP)		0.102	0.138	0.169	0.193	0.231

Note: This table reports the Quadratic Probability Score (QPS) for estimating U.S. business cycle turning points from univariate models using different combination schemes (Bayesian model averaging (BMA), dynamic model averaging (DMA), and an equal-weight scheme for the univariate and bivariate models described in sections 2.1 and 2.2). The first estimation sample extends from February 1960 to December 1978, and it is recursively expanded until the end of the sample is reached (September 2013). Boldface indicates the model with the lowest QPS for a given horizon. Statistically significant reductions in QPS according to the Diebold-Mariano-West test are marked using ***(1% significance level), **(5% significance level) and *(10% significance level).

Table 7: Out-of-sample Quadratic Probability Score - Multivariate models

	Employment					
	Forecast horizon (months)	0	1	2	3	6
Dynamic Model	Likelihood-based	0.317	0.323	0.329	0.330	0.320
Averaging	QPS-based	0.202***	0.214***	0.231***	0.245***	0.269***
alpha=0.99	Combined	0.256**	0.264**	0.271**	0.277**	0.279**
Dynamic Model	Likelihood-based	0.289	0.299	0.309	0.316	0.312
Averaging	QPS-based	0.210***	0.222***	0.237***	0.251***	0.272***
alpha=0.95	Combined	0.240**	0.251**	0.263**	0.271***	0.280***
Bayesian Model	Likelihood-based	0.260	0.265	0.269	0.271	0.267
Averaging	QPS-based	0.231	0.244	0.257	0.269	0.281
	Combined	0.260	0.265	0.269	0.271	0.267
Equal weight		0.224	0.237	0.251	0.264	0.279
	Industrial Production					
Dynamic Model	Likelihood-based	0.150	0.176	0.201	0.219	0.237
Averaging	QPS-based	**0.092****	**0.129****	**0.162***	0.188	0.227
alpha=0.99	Combined	0.096**	0.133**	0.164**	**0.187***	**0.224**
Dynamic Model	Likelihood-based	0.152	0.178	0.202	0.218	0.236
Averaging	QPS-based	0.092**	0.130**	0.163*	0.188	0.227
alpha=0.95	Combined	0.099**	0.136**	0.167*	0.191	0.225
Bayesian Model	Likelihood-based	0.114	0.149	0.180	0.201	0.229
Averaging	QPS-based	0.115	0.150	0.180	0.202	0.230
	Combined	0.113**	0.148*	0.179**	0.200**	0.229
Equal weight		0.109	0.144	0.175	0.198	0.228

Note: This table reports the Quadratic Probability Score (QPS) for estimating U.S. business cycle turning points from multivariate models using different combination schemes (Bayesian model averaging (BMA), dynamic model averaging (DMA), and an equal-weight scheme for the univariate and bivariate models described in sections 2.1 and 2.2. The first estimation sample extends from February 1960 to December 1978, and it is recursively expanded until the end of the sample is reached (September 2013). Boldface indicates the model with the lowest QPS for a given horizon. Statistically significant reductions in QPS according to the Diebold-Mariano-West test are marked using ***(1% significance level), **(5% significance level) and *(10% significance level).

6 Conclusions

This paper provides an extension to the literature on model averaging when one is interested in regime classification. In detail, we modify the standard Bayesian model averaging (BMA) and dynamic model averaging (DMA) combination schemes so as to make the weights depend on past performance to detect regime changes using the quadratic probability score (QPS) to measure the models' ability to classify regimes. The intuition for doing so is relatively straightforward: a model that performs well for continuous forecasts may not necessarily do so for discrete forecasts. Therefore, standard weighting schemes based only on the models' likelihood may not be appropriate in a context of regime classifications.

In an empirical application to forecasting U.S. recessions using state-level employment data, we show the relevance of this framework. In particular, the out-of-sample exercise suggests that weighting schemes based on the QPS outperform weighting schemes based exclusively on the likelihood. In addition, we find that weighting schemes based on the QPS provide timely updates of the U.S. business cycle regimes, in that they typically precede the NBER announcements of business cycle peaks and troughs, and compare favorably with competing models. Also, in both our simulation experiment and empirical application, DMA tends to outperform BMA, suggesting that it is important to allow for time variation in the models' weights.

There are a number of possible extensions of our analysis. First, one could use a broader set of variables in the empirical analysis using for example quarterly GDP growth as a target variable and a broader set of covariates. Mixed-frequency data models could then be used to tackle the mismatch of frequency between the target variable and the covariates. However, doing so would raise complications in terms of computational time since more demanding Bayesian methods would be needed for the estimation of the models. This is likely to prove intractable in a forecasting exercise with a long enough evaluation sample. Second, Wright (2013) emphasizes the importance of seasonal adjustment methods when analyzing U.S. employment data. This is certainly a very important avenue for further work, however it remains unclear the way seasonal adjustment should be performed. We therefore abstracted from this issue, and concentrated our analysis based on the traditional approach of using pre-seasonally adjusted data before estimating models.

References

Belmonte, M. A. G. and Koop, G. (2014). Model Switching and Model Averaging in Time-Varying Parameter Regression Models. *Advances in Econometrics*, forthcoming.

Berge, T. (2013). Predicting recessions with leading indicators: model averaging and selection over the business cycle. *Federal Reserve Bank of Kansas City Working Paper 13-05.*

Berge, T. J. and Jordà, O. (2011). Evaluating the Classification of Economic Activity into Recessions and Expansions. *American Economic Journal: Macroeconomics*, 3(2):246–77.

Billio, M., Casarin, R., Ravazzolo, F., and van Dijk, H. K. (2012). Combination schemes for turning point predictions. *The Quarterly Review of Economics and Finance*, 52(4):402–412.

Billio, M., Casarin, R., Ravazzolo, F., and van Dijk, H. K. (2013). Time-varying combinations of predictive densities using nonlinear filtering. *Journal of Econometrics*, 177(2):213–232.

Cakmakli, C., Paap, R., and van Dijk, D. (2013). Measuring and predicting heterogeneous recessions. *Journal of Economic Dynamics and Control*, 37(11):2195–2216.

Camacho, M., Perez-Quiros, G., and Poncela, P. (2012). Markov-switching dynamic factor models in real time. *CEPR Discussion Paper*, DP8866.

Camacho, M., Perez-Quiros, G., and Poncela, P. (2014). Extracting nonlinear signals from several economic indicators. *Journal of Applied Econometrics*, forthcoming.

Chan, J. C., Koop, G., Leon-Gonzalez, R., and Strachan, R. W. (2012). Time Varying Dimension Models. *Journal of Business & Economic Statistics*, 30(3):358–367.

Chauvet, M. (1998). An Econometric Characterization of Business Cycle Dynamics with Factor Structure and Regime Switching. *International Economic Review*, 39(4):969–96.

Chauvet, M. and Hamilton, J. D. (2006). Dating Business Cycle Turning Points in Real Time. *Nonlinear Time Series Analysis of Business Cycles.*

Chauvet, M. and Piger, J. (2008). A Comparison of the Real-Time Performance of Business Cycle Dating Methods. *Journal of Business Economics and Statistics*, 26(1):42–49.

Chib, S. (1995). Marginal Likelihood From the Gibbs Output. *Journal of the American Statistical Association*, 90:1313–1321.

Croushore, D. and Stark, T. (2001). A real-time data set for macroeconomists. *Journal of Econometrics*, 105(1):111–130.

Diebold, F. X. and Mariano, R. S. (1995). Comparing Predictive Accuracy. *Journal of Business & Economic Statistics*, 13(3):253–63.

Elliott, G. and Timmermann, A. (2005). Optimal Forecast Combination Under Regime Switching. *International Economic Review*, 46(4):1081–1102.

Fruhwirth-Schnatter, S. (2004). Estimating marginal likelihoods for mixture and Markov switching models using bridge sampling techniques. *Econometrics Journal*, 7(1):143–167.

Geweke, J. and Amisano, G. (2011). Optimal prediction pools. *Journal of Econometrics*, 164(1):130–141.

Granger, C. W. J. and Terasvirta, T. (1999). A simple nonlinear time series model with misleading linear properties. *Economics Letters*, 62(2):161–165.

Guérin, P. and Marcellino, M. (2013). Markov-Switching MIDAS Models. *Journal of Business & Economic Statistics*, 31(1):45–56.

Hamilton, J. D. (1989). A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle. *Econometrica*, 57(2):357–84.

Hamilton, J. D. (2011). Calling recessions in real time. *International Journal of Forecasting*, 27(4):1006–1026.

Hamilton, J. D. and Owyang, M. T. (2012). The Propagation of Regional Recessions. *The Review of Economics and Statistics*, 94(4):935–947.

Hamilton, J. D. and Perez-Quiros, G. (1996). What Do the Leading Indicators Lead? *The Journal of Business*, 69(1):27–49.

Hoeting, J. A., Madigan, D., Raftery, A. E., and Volinsky, C. T. (1999). Bayesian Model Averaging: A Tutorial. *Statistical Science*, 14(4):382–417.

Kholodilin, K. A. and Yao, V. W. (2005). Measuring and predicting turning points using a dynamic bi-factor model. *International Journal of Forecasting*, 21(3):525–537.

Kim, C. and Nelson, C. (1999). *State-space models with regime switching: classical and Gibbs-sampling approaches with applications*. The MIT press.

Kim, C.-J. (1994). Dynamic linear models with Markov-switching. *Journal of Econometrics*, 60(1-2):1–22.

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.*, 80(2):188–201.

Kim, M.-J. and Yooa, J.-S. (1995). New index of coincident indicators: A multivariate Markov switching factor model approach. *Journal of Monetary Economics*, 36(3):607–630.

Koop, G. (2003). *Bayesian Econometrics*. John Wiley & Sons Ltd.

Koop, G. (2014). Forecasting with dimension switching VARs. *International Journal of Forecasting*, 30(2):280–290.

Koop, G. and Korobilis, D. (2011). UK macroeconomic forecasting with many predictors: Which models forecast best and when do they do so? *Economic Modelling*, 28(5):2307–2318.

Koop, G. and Korobilis, D. (2012). Forecasting Inflation Using Dynamic Model Averaging. *International Economic Review*, 53(3):867–886.

Koop, G. and Korobilis, D. (2013). Large time-varying parameter VARs. *Journal of Econometrics*, 177(2):185–198.

Koop, G. and Onorante, L. (2012). Estimating Phillips curves in turbulent times using the ECB's survey of professional forecasters. *Working Paper Series European Central Bank*, (1422).

Koop, G. and Onorante, L. (2013). Macroeconomic Nowcasting Using Google Probabilities. *mimeo*.

Koop, G. and Tole, L. (2013). Forecasting the European Carbon Market. *Journal of the Royal Statistical Society Series A*, 176(3):723–741.

Leiva-Leon, D. (2014). A New Approach to Infer Changes in the Synchronization of Business Cycle Phases. *Working Paper Series Bank of Canada*, (2014-38).

Del Negro, M., Giannoni, M. P., and Schorfheide, F. (2013). Dynamic Prediction Pools: An Investigation of Financial Frictions and Forecasting Performance. *mimeo*.

Nalewaik, J. J. (2012). Estimating Probabilities of Recession in Real Time Using GDP and GDI. *Journal of Money, Credit and Banking*, 44(1):235–253.

Newton, M. A. and Raftery, A. E. (1994). Approximate Bayesian inference with the weighted likelihood boostrap. *Journal of the Royal Statistical Society: Series B*, 56:3–26.

Owyang, M. T., Piger, J., and Wall, H. J. (2005). Business Cycle Phases in U.S. States. *The Review of Economics and Statistics*, 87(4):604–616.

Owyang, M. T., Piger, J. M., and Wall, H. J. (2014). Forecasting national recessions using state level data. *Journal of Money, Credit, and Banking*, forthcoming.

Raftery, A., Karny, M., and Ettler, P. (2010). Online Predictions Under Model Uncertainty Via Dynamic Model Averaging: Application to a Cold Rolling Mill. *Technometrics*, 52:52–66.

Raftery, A. E., Madigan, D., and Hoeting, J. A. (1998). Bayesian Model Averaging for Linear Regression Models. *Journal of the American Statistical Association*, 92:179–191.

Timmermann, A. (2006). Forecast Combinations. *Handbook of Economic Forecasting*, 1:135–196.

West, K. D. (1996). Asymptotic Inference about Predictive Ability. *Econometrica*, 64(5):1067–84.

Wright, J. H. (2013). Unseasonal Seasonals? *Brookings Papers on Economic Activity*, 47(2 (Fall)):65–126.

7 Appendix

A Bayesian Parameter Estimation

We follow the multi-move Gibbs-sampling procedure in Kim and Nelson (1999) to estimate the parameters and produce the inference on regimes for the univariate and bivariate Markov-switching models. For brevity we only illustrate the case of the bivariate model, the univariate case being already fully described in Kim and Nelson (1999).

A.1 Priors

For the mean and variance parameters, the Independent Normal-Wishart prior distribution is used:[11]

$$p(\mu, \Sigma^{-1}) = p(\mu)p(\Sigma^{-1}),$$

where

$$\mu \sim N(\underline{\mu}, \underline{V}_\mu), \;\; \Sigma^{-1} \sim W(\underline{S}^{-1}, \underline{\upsilon}),$$

and the associated hyperparameters are $\underline{\mu} = (-1, 2, -1, 2)'$, $\underline{V}_\mu = I$, $\underline{S}^{-1} = I$, $\underline{\upsilon} = 0$.

For the transition probabilities, Beta distributions are used as conjugate priors:

$$p_{k,00} \sim Beta(u_{k,11}, u_{k,10}), \quad p_{k,11} \sim Beta(u_{k,00}, u_{k,01}), \quad \text{for } k = a, b$$

with hyperparameters $u_{k,01} = 2$, $u_{k,00} = 8$, $u_{k,10} = 1$ and $u_{k,11} = 9$ for $k = a, b$.

A.2 Drawing $\tilde{S}_{a,T}$ and $\tilde{S}_{b,T}$ given μ, Σ, $p_{a,00}$, $p_{a,11}$, $p_{b,00}$, $p_{b,11}$, and $\tilde{y}_T$.

To make inference on the dynamics of the state variable $\tilde{S}_{k,T}$, for $k = a, b$, we need to compute draws from the conditional distributions:

$$g(\tilde{S}_{k,T}|\theta, \tilde{y}_T) = g(S_{k,T}|\tilde{y}_T) \prod_{t=1}^{T} g(S_{k,t}|S_{k,t+1}, \tilde{y}_t).$$

To obtain the two terms in the right hand side of the equation above, the following two steps are employed:

Step 1: Run the Hamilton filter to obtain $g(S_{k,t}|\tilde{y}_t)$ for $t = 1, 2, \ldots, T$, and save them. The last iteration, i.e. for $t = T$, provides the first term of the equation.

Step 2: The product in the second term can be obtained for $t = T-1, T-2, \ldots, 1$, with the following result:

$$\begin{aligned} g(S_{k,t}|\tilde{y}_t, S_{k,t+1}) &= \frac{g(S_{k,t}, S_{k,t+1}|\tilde{y}_t)}{g(S_{k,t+1}|\tilde{y}_t)} \\ &\propto g(S_{k,t+1}|S_{k,t})g(S_{k,t}|\tilde{y}_t), \end{aligned}$$

[11] In the case of the univariate model, we use the Normal-Gamma prior distribution.

where $g(S_{k,t+1}|S_{k,t})$ corresponds to the transition probabilities of $S_{k,t}$ and $g(S_{k,t}|\tilde{y}_t)$ were saved in Step 1. Then, it is possible to compute

$$\Pr[S_{k,t}=1|S_{k,t+1},\tilde{y}_t]=\frac{g(S_{k,t+1}|S_{k,t}=1)g(S_{k,t}=1|\tilde{y}_t)}{\sum_{j=0}^{1}g(S_{k,t+1}|S_{k,t}=j)g(S_{k,t}=j|\tilde{y}_t)},$$

and generate a random number from a $U[0,1]$ distribution. If that number is less than or equal to $\Pr[S_{k,t}=1|S_{k,t+1},\tilde{y}_t]$, then $S_{k,t}=1$, otherwise $S_{k,t}=0$.

A.3 Drawing $p_{a,00}$, $p_{a,11}$, $p_{b,00}$ and $p_{b,11}$ given $\tilde{S}_{a,T}$ and $\tilde{S}_{b,T}$.

The likelihood function of $p_{k,00}$, $p_{k,11}$, for $k=a,b$, is given by:

$$L(p_{k,00},p_{k,11}|\tilde{S}_{k,T})=p_{k,00}^{n_{00}}(1-p_{k,00}^{n_{01}})p_{k,11}^{n_{11}}(1-p_{k,11}^{n_{10}}),$$

where $n_{k,ij}$ refers to the transitions from state i to j, accounted for in $\tilde{S}_{k,T}$. Combining the corresponding prior distribution with the likelihood, the posterior distribution reads as

$$p(p_{k,00},p_{k,11}|\tilde{S}_{k,T})\propto p_{k,00}^{u_{k,00}+n_{k,00}-1}(1-p_{k,00})^{u_{k,01}+n_{k,01}-1}p_{k,11}^{u_{k,11}+n_{k,11}-1}(1-p_{k,11})^{u_{k,10}+n_{k,10}-1}$$

which indicates that draws of the transition probabilities will be taken from

$$p_{k,00}|\tilde{S}_{k,T}\sim Beta(u_{k,00}+n_{k,00},u_{k,01}+n_{k,01}),\quad p_{k,11}|\tilde{S}_{k,T}\sim Beta(u_{k,11}+n_{k,11},u_{k,10}+n_{k,10}).$$

A.4 Drawing μ given, Σ, $\tilde{S}_{a,T}$, $\tilde{S}_{b,T}$, and $\tilde{y}_T$.

The bivariate Markov-switching model can be compactly expressed as

$$\begin{aligned}\begin{bmatrix} y_{a,t} \\ y_{b,t}\end{bmatrix} &= \begin{bmatrix} 1 & S_{a,t} & 0 & 0 \\ 0 & 0 & 1 & S_{b,t}\end{bmatrix}\begin{bmatrix}\mu_{a,0}\\ \mu_{a,1}\\ \mu_{b,0}\\ \mu_{b,1}\end{bmatrix}+\begin{bmatrix}\varepsilon_{a,t}\\ \varepsilon_{b,t}\end{bmatrix},\quad \begin{bmatrix}\varepsilon_{a,t}\\ \varepsilon_{b,t}\end{bmatrix}\sim N\left(\begin{bmatrix}0\\0\end{bmatrix},\begin{bmatrix}\sigma_a^2 & \sigma_{ab}\\ \sigma_{ab} & \sigma_b^2\end{bmatrix}\right)\\ y_t &= \bar{S}_t\mu+\xi_t,\quad \xi_t\sim N(\mathbf{0},\Sigma),\end{aligned}$$

stacking as:

$$y=\begin{bmatrix}y_1\\ y_2\\ \vdots\\ y_T\end{bmatrix},\ \bar{S}=\begin{bmatrix}\bar{S}_1\\ \bar{S}_2\\ \vdots\\ \bar{S}_T\end{bmatrix},\ \text{and } \xi=\begin{bmatrix}\xi_1\\ \xi_2\\ \vdots\\ \xi_T\end{bmatrix},$$

the model remains written as a normal linear regression with an error covariance matrix of a particular form:

$$y=S\mu+\xi,\quad \xi\sim N(\mathbf{0},I\otimes\Sigma)$$

Using the corresponding likelihood function, the conditional posterior distribution for the intercepts reads as:

$$\mu|\tilde{S}_{a,T}, \tilde{S}_{b,T}, \Sigma^{-1}, \tilde{y}_T \sim N(\overline{\mu}, \overline{V}_\mu),$$

where

$$\begin{aligned}
\overline{V}_\mu &= \left(\underline{V}_\mu^{-1} + \sum_{t=1}^{T} \bar{S}_t' \Sigma^{-1} \bar{S}_t \right)^{-1} \\
\overline{\mu} &= \overline{V}_\mu \left(\underline{V}_\mu^{-1} \underline{\mu} + \sum_{t=1}^{T} \bar{S}_t' \Sigma^{-1} y_t \right).
\end{aligned}$$

When drawing $\mu = (\mu_{a,0}, \mu_{a,1}, \mu_{b,0}, \mu_{b,1})'$, we impose the constraint that $\mu_{a,1} > 0$ and $\mu_{b,1} > 0$ to ensure identification of the regimes in the model.

A.5 Drawing Σ given μ, $\tilde{S}_{a,T}$, $\tilde{S}_{b,T}$, and $\tilde{y}_T$.

Conditional on the mean, state variables and the data, the conditional posterior distribution for the variance-covariance matrix parameters reads as:

$$\Sigma^{-1}|\tilde{S}_{a,T}, \tilde{S}_{b,T}, \mu, \tilde{y}_T \sim W(\overline{S}^{-1}, \overline{\upsilon}),$$

$$\begin{aligned}
\overline{\upsilon} &= T + \underline{\upsilon} \\
\overline{S} &= \underline{S} + \sum_{t=1}^{T} \left(y_t - \bar{S}_t \mu \right) \left(y_t - \bar{S}_t \mu \right)',
\end{aligned}$$

after Σ^{-1} is generated, the elements in Σ are recovered.

The above steps are iterated 7000 times, discarding the first 2000 iterations to mitigate the effect of the initial conditions.

B Dynamic Model Averaging in Markov-Switching Models

To compute the time-varying weights associated to each Markov-switching model, we follow an algorithm that combines the Hamilton filter with the prediction and updating equations used in the dynamic model averaging approach in Raftery et al. (2010).

At any given period t, we compute the following steps for all the models under consideration:

- **Step 1:** Using the corresponding transition probabilities $p(S_t^k|S_{t-1}^k)$, compute the predicted regime probabilities for any given model k given past information ψ_{t-1}, $P(S_t^k|M_t = k, \psi_{t-1})$.[12]

$$P(S_t^k, S_{t-1}^k|\psi_{t-1}) = p(S_t^k|S_{t-1}^k)P(S_{t-1}^k|\psi_{t-1}) \tag{30}$$

$$p(S_t^k|M_t = k, \psi_{t-1}) = \sum_{S_{t-1}^k} P(S_t^k, S_{t-1}^k|\psi_{t-1}). \tag{31}$$

  Then, the marginal likelihood is calculated from the predicted probabilities:

$$f_k(y_t|\psi_{t-1}) = \sum_{S_t^k}\sum_{S_{t-1}^k} f_k(y_t|S_t^k, S_{t-1}^k, \psi_{t-1})P(S_t^k, S_{t-1}^k|\psi_{t-1}) \tag{32}$$

- **Step 2:** Let $\pi_{t|t-1,k} = P(M_t = k|\psi_{t-1})$ be the predictive probability associated with the k-th Markov-switching model at time t given the information up to $t-1$. Starting with an equal weight initial model probability $P(M_0)$, we consider three different approaches to compute the model updated probability:

  A) *Likelihood-based approach:* we follow the updating criterion of Raftery et al. (2010), which uses the marginal likelihood:

$$\pi_{t|t,k} = \frac{\pi_{t|t-1,k} f_k(y_t|\psi_{t-1})}{\sum_{j=1}^K \pi_{t|t-1,j} f_j(y_t|\psi_{t-1})}. \tag{33}$$

  B) *Combined approach:* we propose to use both types of information, i.e. the inverse cumulative QPS and the marginal likelihood:

$$\pi_{t|t,k} = \frac{\pi_{t|t-1,k} f_k(y_t|\psi_{t-1})/Q_{k,t|t}}{\sum_{j=1}^K \pi_{t|t-1,j} f_j(y_t|\psi_{t-1})/Q_{j,t|t}}, \tag{34}$$

  where $Q_{k,t|t} = \frac{2}{t}\sum_{\tau=1}^t (P(S_\tau^k = 0|M_t = k, \psi_t) - NBER_\tau)^2$.

[12] The Hamilton filter is initialized with the ergodic probabilities $P(S_0)$.

C) *QPS-based approach:* we also propose to use just the inverse of the cumulative QPS instead of the likelihood, since the it focuses only on ability of models to detect regime shifts:

$$\pi_{t|t,k} = \frac{\pi_{t|t-1,k} Q^{-1}_{k,t|t}}{\sum_{j=1}^{K} \pi_{t|t-1,j} Q^{-1}_{j,t|t}}. \tag{35}$$

- **Step 3:** Use the marginal likelihood, $f_k(y_t|\psi_{t-1})$, to compute the updated regime probabilities for any given model k, $P(S_t^k|M_t = k, \psi_t)$, as follows:

$$\begin{aligned} P(S_t^k, S_{t-1}^k|\psi_t) &= \frac{f_k(y_t, S_t^k, S_{t-1}^k|\psi_{t-1})}{f_k(y_t|\psi_{t-1})} \\ &= \frac{f_k(y_t|S_t^k, S_{t-1}^k, \psi_{t-1}) P(S_t^k, S_{t-1}^k|\psi_{t-1})}{f_k(y_t|\psi_{t-1})} \end{aligned} \tag{36}$$

$$P(S_t^k|M_t = k, \psi_t) = \sum_{S_{t-1}^k} P(S_t^k, S_{t-1}^k|\psi_t), \tag{37}$$

which are used in Step 1 of the next iteration .

- **Step 4:** Compute the predicted probability associated to the k-th model, $\pi_{t+1|t,k}$, by relying on Raftery et al. (2010) and using the forgetting factor α, as follows:

$$\pi_{t+1|t,k} = \frac{\pi^{\alpha}_{t|t,k}}{\sum_{j=1}^{K} \pi^{\alpha}_{t|t,j}}, \tag{38}$$

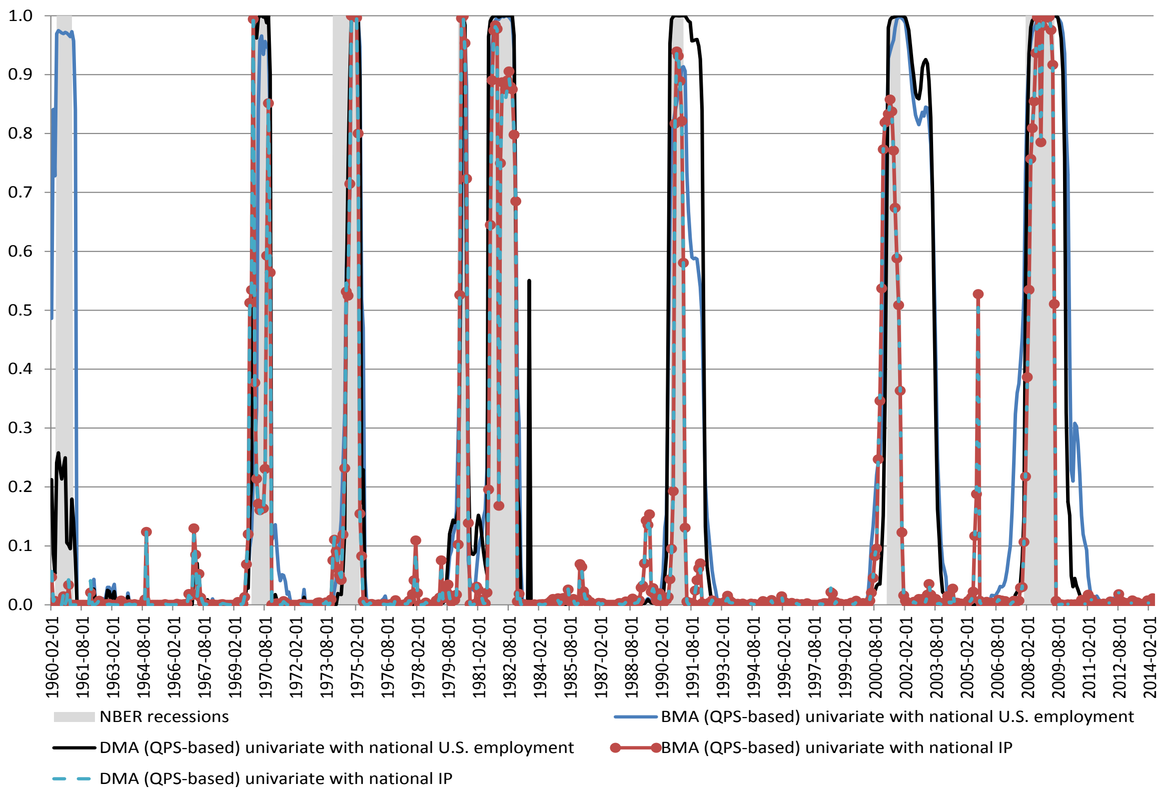
which are used in Step 2 of the next iteration .

We repeat the steps above for each model at each period of time $t = 1, ..., T$. The output of the algorithm consists of the regimes probabilities for each model, $P(S_t^k|M_t = k, \psi_t)$, and the model probabilities for each time period, $\pi_{t|t,k} = P(M_t = k|\psi_t)$. Therefore, we compute the expected regime probabilities by averaging them across models:

$$P(\mathbf{S}_t|\psi_t) = \sum_{k=1}^{K} P(S_t^k|M_t = k, \psi_t) P(M_t = k|\psi_t). \tag{39}$$

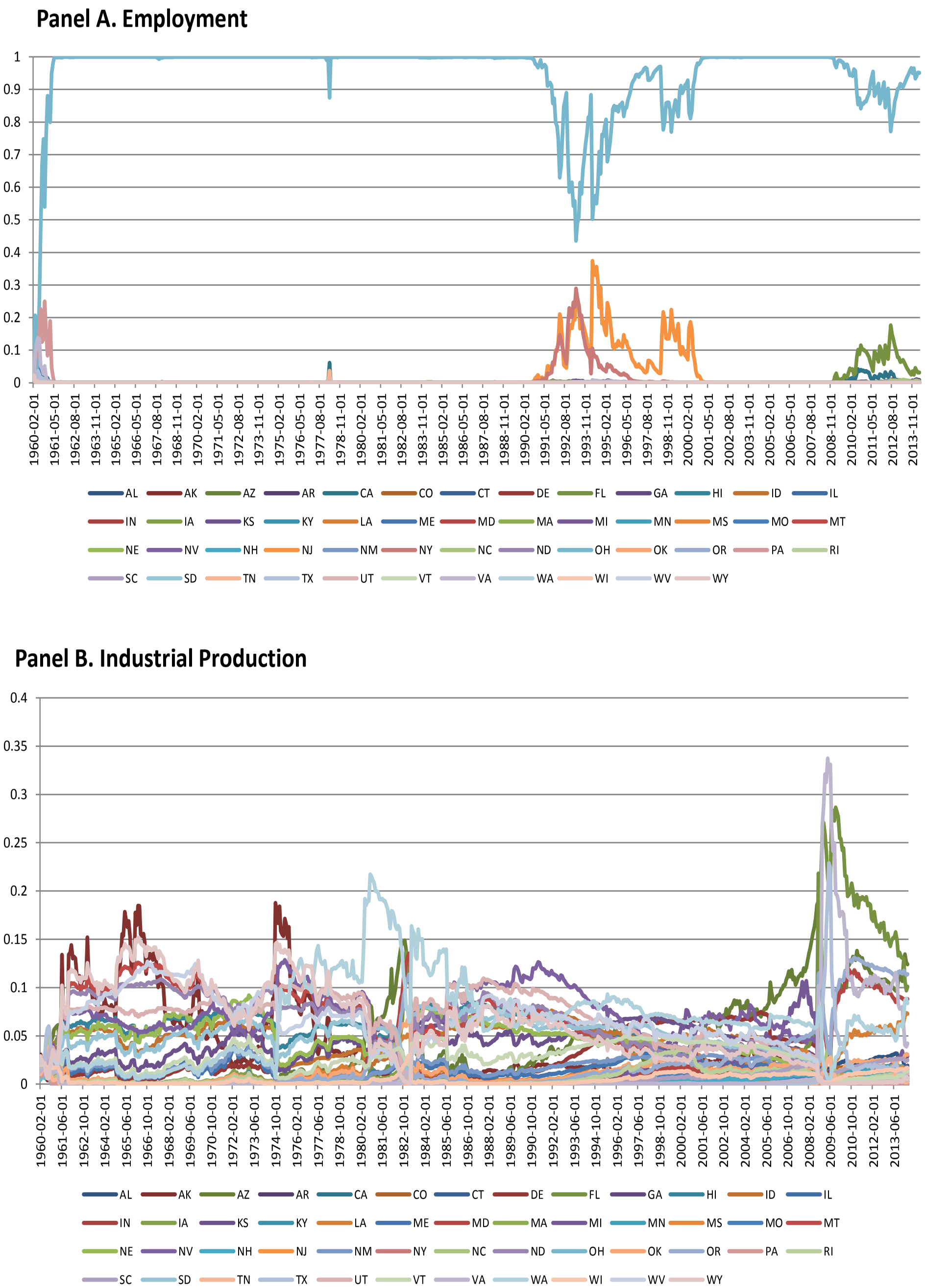
The aggregated probability $P(\mathbf{S}_t|\psi_t)$ from the above equation is used to assess the performance of all models using the dynamic model averaging approach.

Figure 1: In-sample Probability of Recession



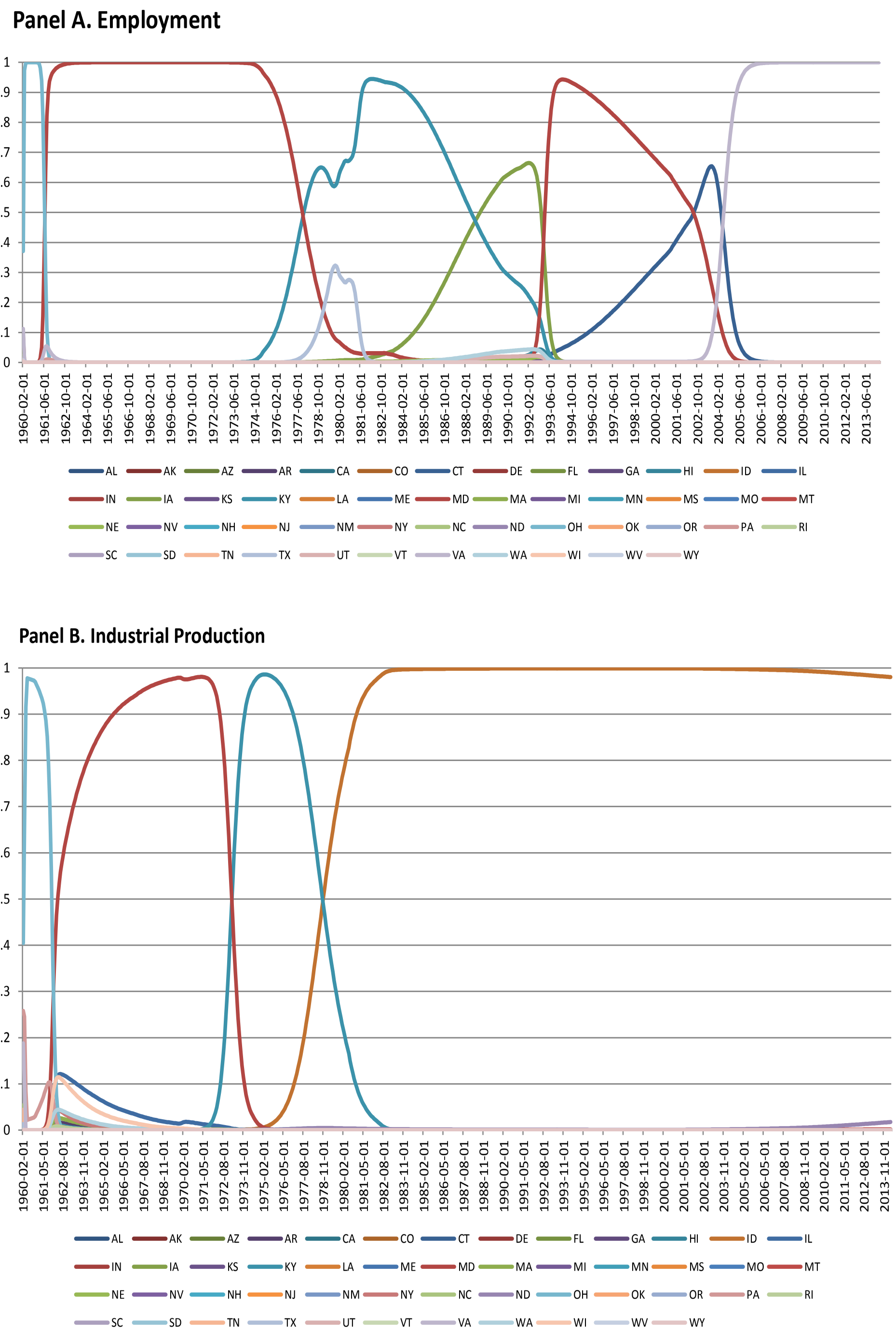
Note: This figure reports the monthly in-sample probability of a recession extending from February 1960 to April 2014 obtained from averaging the results from individual models using different combination schemes (BMA weights based on the QPS and DMA weights based on the QPS).

Figure 2: In-sample model weights from dynamic model averaging (likelihood-based)



Note: This figure reports the weights obtained when averaging the results from univariate models using DMA weights based on the likelihood (with forgetting factor $\alpha = 0.99$). Panel A shows the results when using employment as a measure of national economic activity and panel B when using industrial production as a measure of national economic activity.

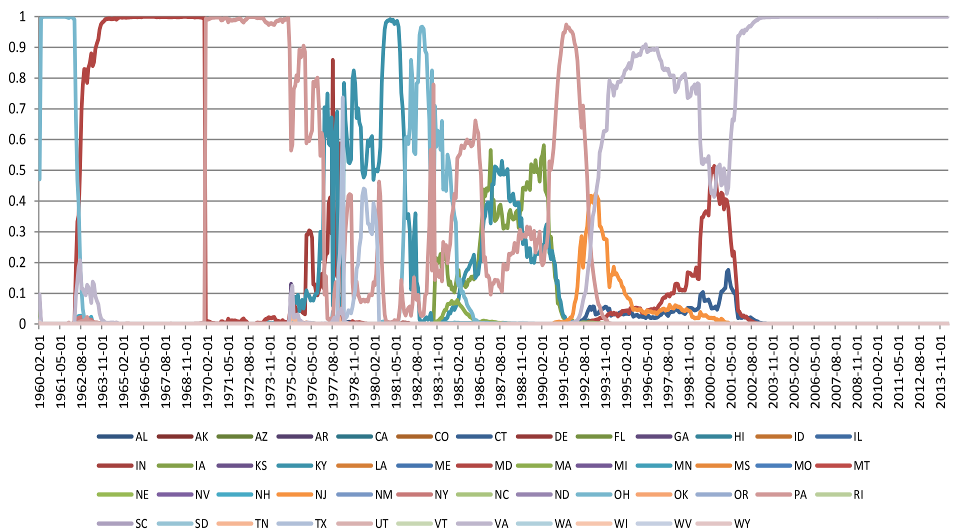
Figure 3: In-sample model weights from dynamic model averaging (QPS-based)



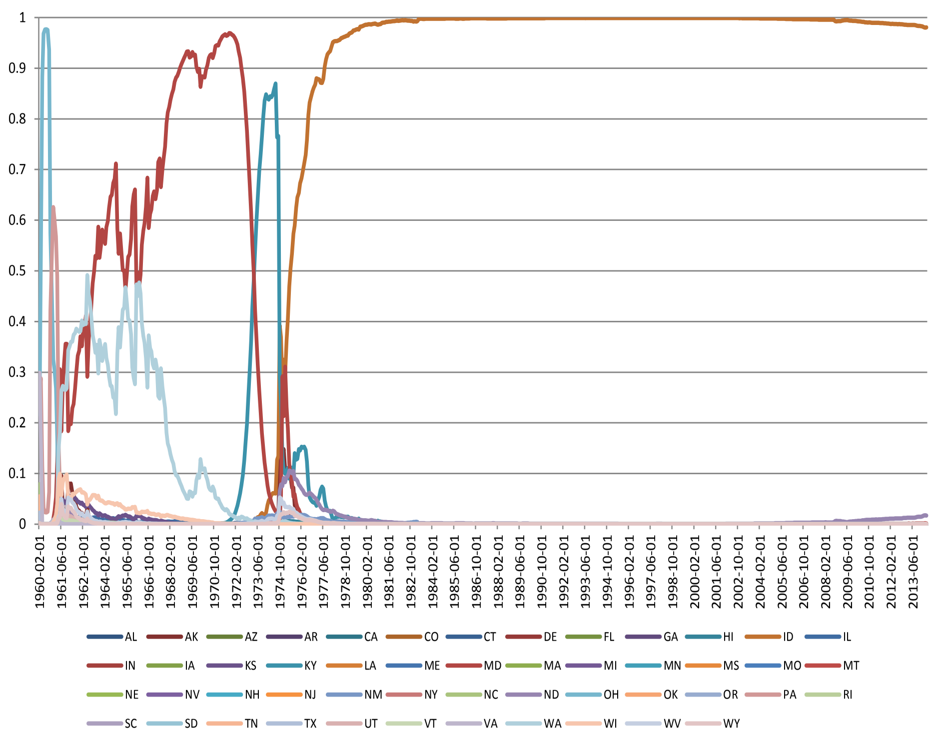
Note: This figure reports the weights obtained when averaging the results from univariate models using DMA weights based on the QPS (with forgetting factor $\alpha = 0.99$). Panel A shows the results when using employment as a measure of national economic activity and panel B when using industrial production as a measure of national economic activity.

Figure 4: In-sample model weights from dynamic model averaging (Combined)

Panel A. Employment

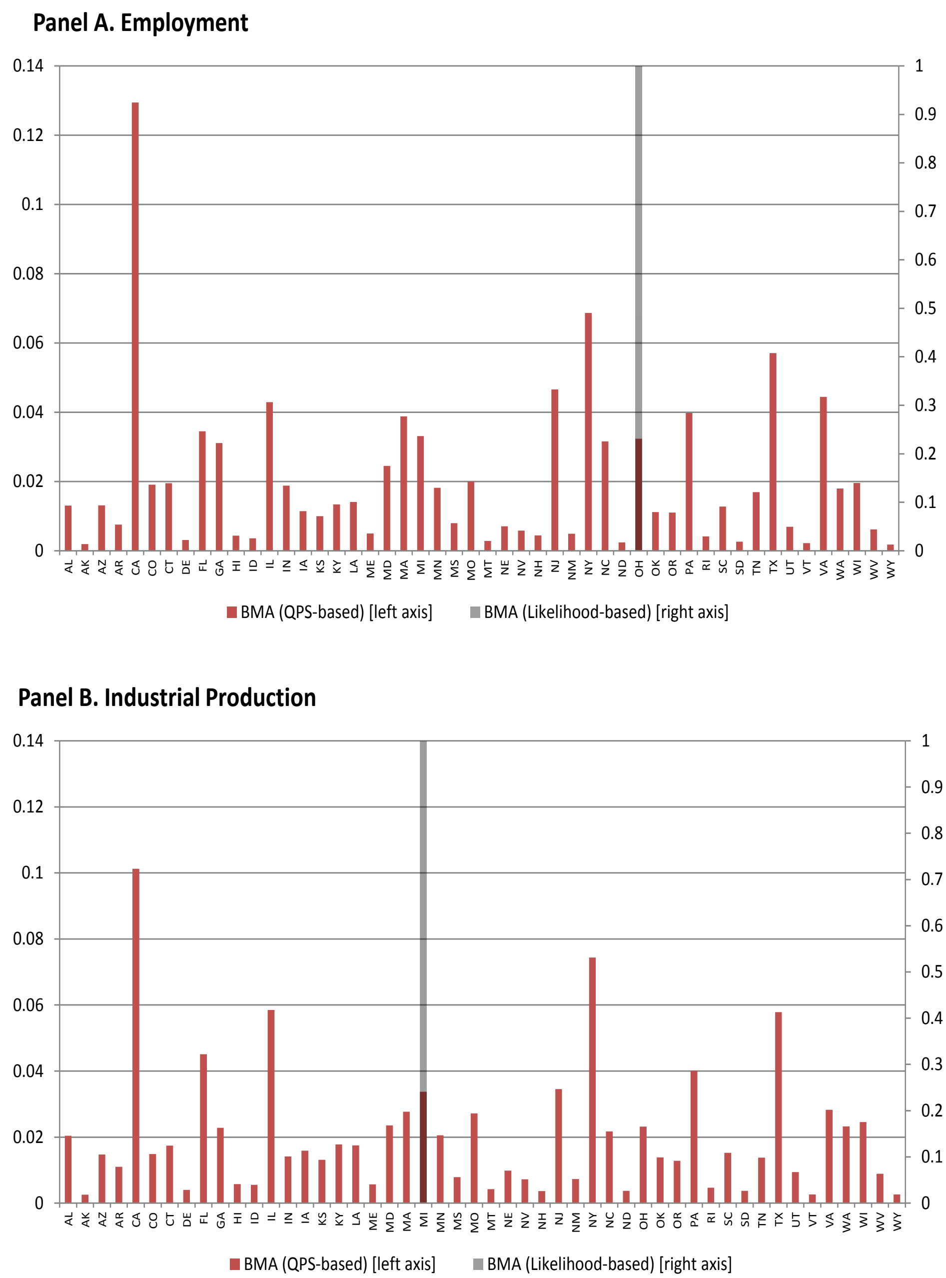


Panel B. Industrial Production



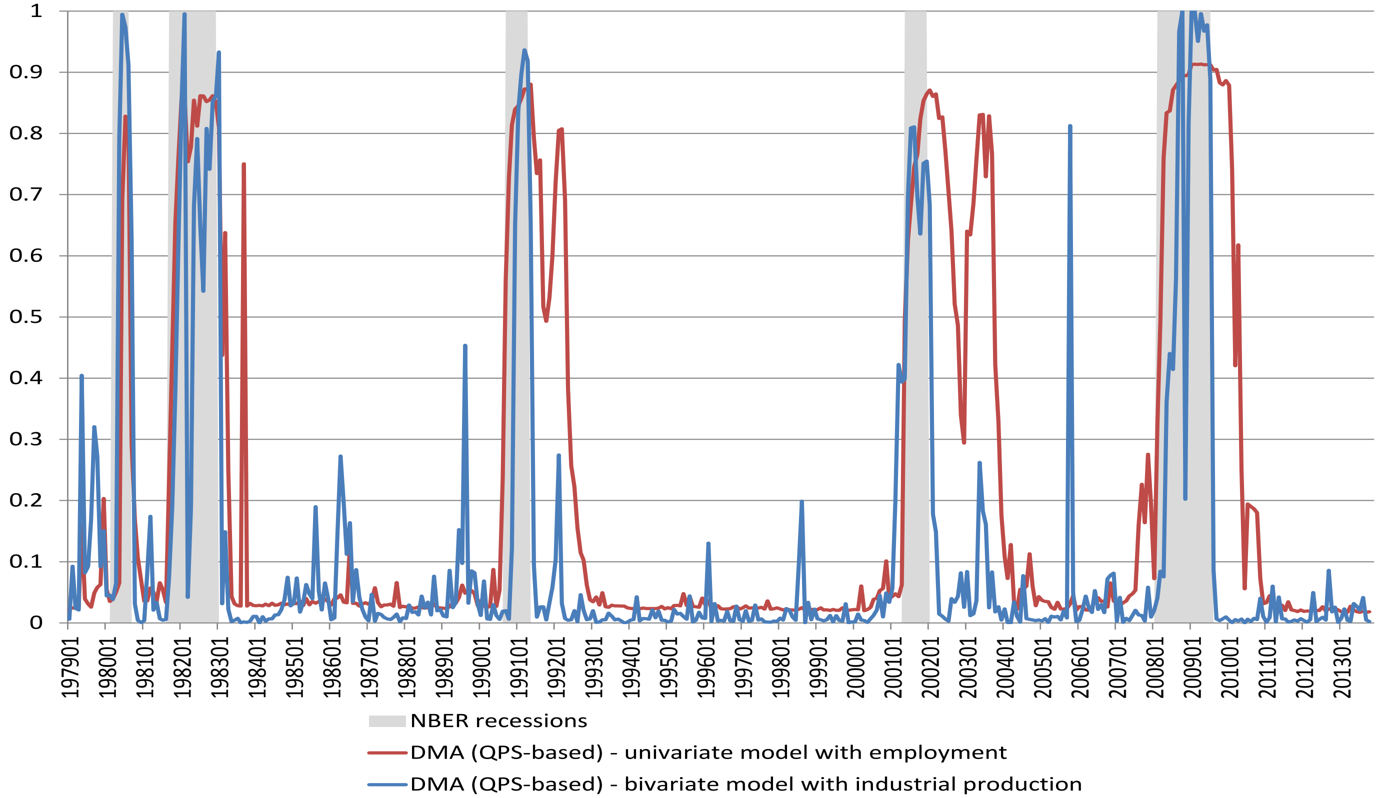
Note: This figure reports the weights obtained when averaging the results from univariate models using DMA weights based on a combination of QPS and predictive likelihood (with forgetting factor $\alpha = 0.99$). Panel A shows the results when using employment as a measure of national economic activity and panel B when using industrial production as a measure of national economic activity.

Figure 5: In-sample model weights from Bayesian model averaging



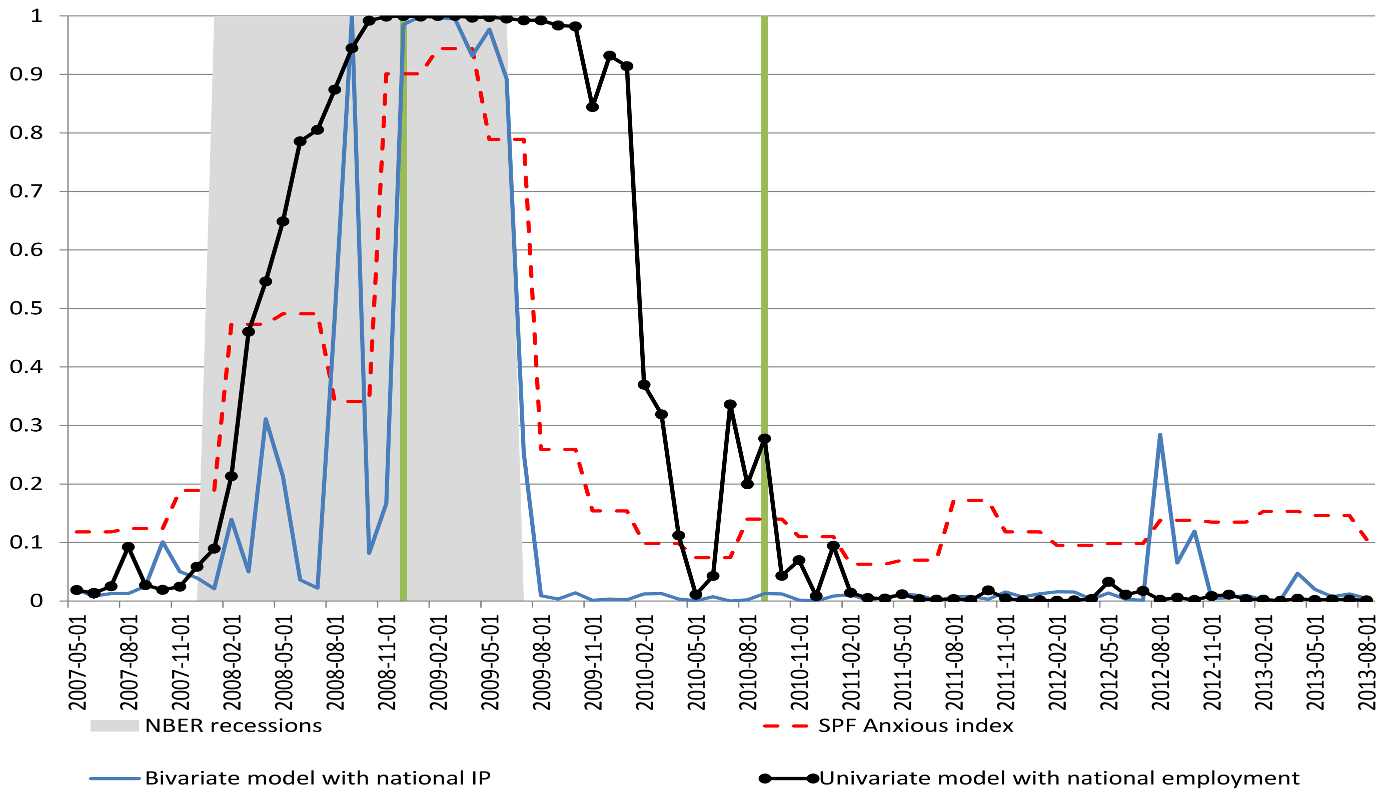
Note: This figure reports the weights obtained when averaging the results from univariate models using BMA weights based on the QPS and marginal likelihood. Panel A shows the results when using employment as a measure of national economic activity and panel B when using industrial production as a measure of national economic activity.

Figure 6: Predicted probability of a U.S. recession (1-month-ahead forecast)



Note: This figure displays the predicted one-month-ahead probability of a recession using the QPS-based DMA combination scheme with selected models.

Figure 7: Real-time probability of a U.S. recession



Note: This figure displays the real-time probability of a recession using the QPS-based DMA combination scheme (univariate model with employment and bivariate model with industrial production) along with the probability of recession derived from the Survey of Professional Forecasters (SPF). The vertical bars indicate the dates of the NBER announcements of the business cycle peak and trough.