BANK OF CANADA
BANQUE DU CANADA

Working Paper/Document de travail
2015-24

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

by Pierre Guérin and Danilo Leiva-Leon

Bank of Canada Working Paper 2015-24

June 2015

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

**by**

**Pierre Guérin[1] and Danilo Leiva-Leon[2]**

[1]International Economic Analysis Department
Bank of Canada
Ottawa, Ontario, Canada K1A 0G9
pguerin@bankofcanada.ca

[2]Banco Central de Chile
dleiva@bcentral.cl



ISSN 1701-9397 

Acknowledgements

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, Robert Lavigne, Michael Ehrmann and seminar participants at the Applied Time Series Econometrics Workshop held at the Federal Reserve Bank of St. Louis, the 2014 CFE conference in Pisa, Bank of Canada and University of California at Riverside for helpful discussions and comments.

## Abstract

This paper introduces new weighting schemes for model averaging when one is interested in combining discrete forecasts from competing Markov-switching models. In particular, we extend two existing classes of combination schemes – Bayesian (static) model averaging and dynamic model averaging – so as to explicitly reflect the objective of forecasting a discrete outcome. Both simulation and empirical exercises show that our new combination schemes outperform competing combination schemes in terms of forecasting accuracy. In the empirical application, we estimate and forecast U.S. business cycle turning points with state-level employment data. We find that forecasts obtained with our best combination scheme provide timely updates of the U.S. business cycles.

*JEL classification: C53, E32, E37*
*Bank classification: Business fluctuations and cycles; Econometric and statistical methods*

## Résumé

Les auteurs présentent de nouvelles méthodes de pondération pour la combinaison de prévisions de variables discrètes issues de différents modèles de Markov à changement de régime. Plus particulièrement, ils étendent deux classes existantes de méthodes de combinaison – combinaison de prévisions établies au moyen de modèles bayésiens (statiques) et combinaison dynamique de prévisions – de manière à correspondre explicitement à l'objectif assigné à l'exercice de prévision d'une variable discrète. Les simulations et l'application empirique montrent qu'en ce qui a trait à l'exactitude des prévisions, les nouvelles méthodes de combinaison surclassent les méthodes de combinaison concurrentes. S'agissant de l'application empirique, les auteurs estiment et prévoient les points de retournement du cycle de l'économie américaine à l'aide de données sur l'emploi provenant des États. Ils constatent que les prévisions obtenues à partir de leur meilleure méthode de combinaison fournissent des renseignements en temps opportun sur les cycles économiques aux États-Unis.

*Classification JEL : C53, E32, E37*
*Classification de la Banque : Cycles et fluctuations économiques; Méthodes économétriques et statistiques*

**Non-Technical Summary**

Combining forecasts from different models is typically perceived as a useful way to mitigate time instability in the forecasting performance of individual models, and thereby ultimately improve the overall forecasting performance. Moreover, regime-switching models are widely used to model non-linear relations in macroeconomics and finance. A key advantage of regime-switching models is their ability to estimate both discrete and continuous outcomes. For example, one may be interested not only in forecasting the level of GDP growth, but also in evaluating whether an economy is heading into recession. It is therefore of natural interest to study how to combine information from competing regime-switching models in a forecasting environment.

A key contribution of this paper is to suggest new combination schemes in the context of regime-switching models, and compare them with existing weighting schemes. In detail, we consider both constant and time-varying weights when calculating the models' weights (i.e., Bayesian model averaging and dynamic model averaging, respectively). Rather than calculating weights exclusively based on how likely they are to best fit the data, we suggest using weights that rely on how well a model describes the underlying regimes. Another contribution of this paper is that it describes how to perform dynamic model averaging when estimating regime-switching models. One key conclusion of this paper is that combination schemes based on their past performance in evaluating regimes perform best when one is estimating and forecasting discrete outcomes.

We first perform a Monte Carlo experiment that relies on simulated data to study in a controlled experiment the different combination schemes outlined in this paper. Second, our empirical application concentrates on predicting national U.S. recessions with state-level employment data using, alternatively, industrial production and employment as a measure of national economic activity. We show that our preferred specifications provide timely updates of the U.S. business cycles. This holds true over an extended evaluation sample for the forecasting exercise that permits us to alleviate concerns about spurious forecasting results, and also over a sample that considers real-time data so as to mimic as closely as possible the information set available at the time the forecasts were made. Overall, we trust that the methods outlined in this paper can be implemented in a large class of applications in applied macroeconomics and finance.

# 1 Introduction

Combining forecasts has frequently been found to produce better forecasts on average than forecasts obtained from the best ex-ante individual forecasting model. This is especially true when there is substantial time instability in the forecasting ability of predictors and when parameter estimation error is sizable (see, e.g., Timmermann (2006)). While there is a large literature on model averaging for linear models, little work has been done to study model-averaging schemes for Markov-switching models. This is a highly relevant issue to study, since Markov-switching models are widely used in macroeconomics and finance. For example, starting from Hamilton (1989), there is a large literature on the estimation and prediction of recessions using regime-switching models. In this paper, we introduce new weighting schemes to combine discrete forecasts from competing Markov-switching models. We show their relevance based on Monte Carlo experiments and an empirical application to predict U.S. recessions with a large set of Markov-switching models.

In the context of linear regressions, an approach increasingly used in empirical studies is Bayesian model averaging (BMA), proposed by Raftery et al. (1998) and Hoeting et al. (1999). This approach is often referred to as static model averaging in that the estimated models' weights are constant over time. Recently, Raftery et al. (2010) proposed the dynamic model averaging (DMA) approach, where the models' weights evolve over time. DMA has been applied to a variety of situations, most notably to forecasting inflation in the context of time-varying parameter (TVP) regression models (see Koop and Korobilis (2012)). In a multivariate context, DMA has been applied to linear vector autoregressive (VAR) models (Koop (2014)) and large TVP-VAR models (Koop and Korobilis (2013)) to forecast inflation, real output and interest rates. An alternative approach for model combination is provided in Elliott and Timmermann (2005), who suggest using 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 the predictive score. Del Negro et al. (2013) provide 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 model-averaging schemes in the context of discrete forecasts. To the best of our knowledge, there are very few related works in this research area. Among the notable exceptions are Billio et al. (2012), who compare the performance of combination schemes for linear and regime-switching models, and Billio et al. (2013), who propose a time-varying combination approach for multivariate predictive densities. Moreover, Berge (2013) compares model selection schemes based on boosting algorithms with a BMA weighting scheme for predicting U.S. recessions based on logistic regressions using a set of economic and financial indicators.

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 Yoo (1995), and Kim and Nelson (1998). Moreover, Chauvet (1998) finds that this type of model performs well at dating 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 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 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 with regime-switching models. He concludes that forecasting gains are likely to be made when averaging inferences from competing models.

This paper contributes to the literature along two dimensions. First, we show how to perform DMA with Markov-switching models. This extension provides a flexible framework that evaluates, at every period of time, the performance of different Markov-switching models to infer the regimes of a target variable. Second, we introduce new weighting schemes for model averaging when the variable to forecast is a discrete outcome. Specifically, we propose models' weights that depend on the past predictive ability of a given model to estimate discrete outcomes. It is intuitive to do so in that a model that performs well for continuous forecasts may not necessarily be helpful for discrete forecasts. Hence, standard weighting schemes that exclusively rely on the likelihood as a measure of model fit may not necessarily be appropriate when forecasting discrete outcomes. When presenting our new models' weights, we consider two classes of combination schemes, static weights (BMA) and time-varying weights (DMA).

We compare the predictive ability of the likelihood-based combination schemes (i.e., the standard approach adopted in the literature) with the predictive ability of our proposed combination schemes. This comparison is performed with simulated data, based on Monte Carlo experiments, and with an empirical application. The empirical application concentrates on predicting U.S. national recessions using state-level employment data. There are several reasons for choosing this application. First, previous studies suggest that employment of specific U.S. states may lead the national business cycle phases in particular episodes (Owyang et al. (2005)). As a result, it is natural to think of the best way to combine information from the different U.S. states to predict a national aggregate. Second, 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 a significant heterogeneity in their cyclical fluctuations, which suggests the need for appropriate combination schemes when analyzing U.S. national and state-level data. Third, Owyang et al. (2014) use state-level data to forecast U.S. recessions with 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 U.S. business cycle phases.[1]

[1]We follow Owyang et al. (2014) in that we also use state-level employment data to predict national

Our main results can be summarized as follows. First, in both our Monte Carlo and empirical experiments, 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 predictive ability to fit discrete outcomes typically outperform combination schemes based on the likelihood only. This is especially true in an out-of-sample context. Second, on average, the best combination scheme in terms of predictive accuracy is obtained with the DMA framework where the weights depend on the past predictive ability to estimate discrete outcomes. Third, the use of regional data improves the forecasting performance compared with models using exclusively national data. Fourth, out-of-sample forecasts obtained with the best combination scheme outperform the anxious index from the Survey of Professional Forecasters for short-term forecasts, which emphasizes the relevance of our framework. In addition, in a purely real-time environment, we also find that our best combination scheme provides timely estimates of the latest U.S. recession.

The paper is organized as follows. Section 2 describes the underlying models we use in the forecasting combination exercise. Section 3 presents the different combination schemes, and details the extensions to the standard combination schemes we implement. 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

For simplicity, in the sequel we adopt a notation that is consistent with our empirical application, but note that the framework we present is general enough to accommodate a large class of applications in macroeconomics and finance.

## 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 the dependent variable and $x_t^k$ denotes a given regressor $k$. The error term, denoted by $v_t^k$, is assumed to be normally distributed, that is, $v_t^k \sim N(0, \sigma_k^2)$. Note also that equation (1) could easily accommodate a set of regressors $X_t^k$ instead of a single predictor $x_t^k$. $S_t^k$ is a standard Markov chain defined by the following constant transition probability:

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

U.S. recessions. However, we focus on regime-switching models rather than probit models that include the NBER dating of business cycle regimes as a dependent variable, the latter approach being problematic in a forecasting context given the substantial publication delay in the announcements of the NBER business cycle turning points.

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

where $M$ is the number of regimes.

In relation to the empirical application, described in Section 5, $y_t$ is the U.S. national employment, while the $x_t^k$'s represent employment at the state level, both in growth rates. Note that this specification differs from the baseline specification in Owyang et al. (2005), since they estimate a 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 Bivariate model

We next 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. $S_t^y$ and $S_t^k$ are two independent Markov chains, and $\epsilon_t^k$ is normally distributed, that is, $\epsilon_t^k \sim N(0, \Sigma_k)$ where $\Sigma_k$ is defined as

$$\Sigma_k = \begin{pmatrix} \sigma_k^{yy} & \sigma_k^{yx} \\ \sigma_k^{yx} & \sigma_k^{xx} \end{pmatrix}.$$

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 bivariate model, assuming that they are independently generated. This implies that regime changes at the national and state level do not necessarily coincide (but are not restricted to differ either). 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 (static) model averaging and dynamic model averaging. This section details the extensions we introduce to these two combination schemes so as to explicitly reflect the objective of estimating and forecasting a discrete outcome when using Markov-switching models.

## 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 bivariate), collected in the vector $\Theta_k$. 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.[2] 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 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, $P(S_t|y_t)$.[3] 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

[2]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.

[3]For ease of exposition, here, we present the case of only 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.

to evaluate discrete outcomes in the business cycle literature.[4] The QPS associated with 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 $t$, $\psi_t$, and $NBER_t$ is a dummy variable that takes 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. Accordingly, the posterior model probability for model $k$ reads as

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

One could use the U.S. employment share of each state as the 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 $M_k$ to explain 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:

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

[4]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 by which to evaluate discrete outcomes.

## 3.2 Dynamic model averaging

### 3.2.1 DMA for Markov-switching models

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 calculate the time-varying weights associated with each Markov-switching model, we suggest the following algorithm that combines the Hamilton filter with the prediction and updating equations used in the DMA approach from 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 $\psi_{t-1}$, $P(S_t^k|M_t = k, \psi_{t-1})$:[5]

$$P(S_t^k, S_{t-1}^k|\psi_{t-1}) = p(S_t^k|S_{t-1}^k)P(S_{t-1}^k|\psi_{t-1}) \tag{19}$$

$$p(S_t^k|M_t = k, \psi_{t-1}) = \sum_{S_{t-1}^k} P(S_t^k, S_{t-1}^k|\psi_{t-1}). \tag{20}$$

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

$$f_k(y_t|\psi_{t-1}) = \sum_{S_t^k}\sum_{S_{t-1}^k} f_k(y_t|S_t^k, S_{t-1}^k, \psi_{t-1})P(S_t^k, S_{t-1}^k|\psi_{t-1}). \tag{21}$$

**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 follow the updating criterion of Raftery et al. (2010), which is based on a measure of model fit for $y_t$, that is, the predictive likelihood:

$$\pi_{t|t,k} = \frac{\pi_{t|t-1,k} f_k(y_t|\psi_{t-1})}{\sum_{j=1}^{K} \pi_{t|t-1,j} f_j(y_t|\psi_{t-1})}. \tag{22}$$

**Step 3:** Use the predictive 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:

$$\begin{aligned} P(S_t^k, S_{t-1}^k|\psi_t) &= \frac{f_k(y_t, S_t^k, S_{t-1}^k|\psi_{t-1})}{f_k(y_t|\psi_{t-1})} \\ &= \frac{f_k(y_t|S_t^k, S_{t-1}^k, \psi_{t-1})P(S_t^k, S_{t-1}^k|\psi_{t-1})}{f_k(y_t|\psi_{t-1})}, \end{aligned} \tag{23}$$

$$P(S_t^k|M_t = k, \psi_t) = \sum_{S_{t-1}^k} P(S_t^k, S_{t-1}^k|\psi_t), \tag{24}$$

which are used in Step 1 of the next iteration.

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

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

$$\pi_{t+1|t,k} = \frac{\pi_{t|t,k}^{\alpha}}{\sum_{j=1}^{K} \pi_{t|t,j}^{\alpha}}, \tag{25}$$

which are used in Step 2 of the next iteration. The forgetting factor $\alpha$ is the coefficient that governs the amount of persistence in the models' weights, and it is set to a fixed value slightly less than one. The higher the $\alpha$, the higher the weight attached to past predictive performance is.[6]

We repeat the steps above for each model at each period of time $t = 1, ..., T$. The output of the algorithm consists of the regime 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:

$$P(\mathbf{S}_t|\psi_t) = \sum_{k=1}^{K} P(S_t^k|M_t = k, \psi_t)P(M_t = k|\psi_t). \tag{26}$$

The average probability $P(\mathbf{S}_t|\psi_t)$ is used to assess the performance of the Markov-switching DMA combination scheme. Notice that DMA also differs from BMA in that no simulation is required to calculate the time-varying models' weights.

### 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 indicate the model's ability to predict business cycle phases. Therefore, the updating equation is replaced by

$$\pi_{t|t,k} = \frac{\pi_{t|t-1,k}\eta_{t|t-1,k}}{\sum_{j=1}^{K} \pi_{t|t-1,j}\eta_{t|t-1,j}}, \tag{27}$$

where

$$\eta_{t|t-1,k} = \frac{f_k(y_t|\psi_{t-1})}{Q_{t|t,k}}, \tag{28}$$

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

$$Q_{t|t,k} = \frac{2}{t}(\sum_{\tau=1}^{t} P(S_\tau^k = 0|\psi_\tau) - NBER_\tau)^2. \tag{29}$$

The model prediction equation remains the same as in equation (25).

[6]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).

### 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 the 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

$$\pi_{t|t,k} = \frac{\pi_{t|t-1,k} Q^{-1}_{t|t,k}}{\sum_{j=1}^{K} \pi_{t|t-1,j} Q^{-1}_{t|t,j}}. \tag{30}$$

# 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 (DGP) that closely mimics the empirical application of the paper, albeit for computational reasons the sample size and the number of predictors are restricted. Equations (31) to (33) detail the DGP. First, the dependent variable $y_t$ is generated according to the following equation:

$$y_t = \mu_0^y + \mu_1^y S_t + \varepsilon_t^y, \tag{31}$$

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:

$$x_{k,t} = \mu_0^k + \mu_1^k S_t + \sigma_k \varepsilon_t^k, \text{ for } k = \{1, ..., K\}, \tag{32}$$

where $\varepsilon_t^k \sim N(0, \sigma_k^2)$.[7]

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

$$\mu_j^k = \mu_j^y + \mu_j^y \epsilon_{k,j}, \text{ for } j = \{0, 1\}, \tag{33}$$

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 $\mu_0^y$ and $\mu_1^y$ 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

[7]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.

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.[8] For ease of computation, we also assume that a single Markov chain $S_t$ drives the changes for both the $y_t$ and $x_t$'s variables.[9]

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 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 $\sigma$), 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, regime 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 United States. 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 (note that state-level data are not available for monthly industrial production). 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 available only since June 2007 from the "Alfred" real-time database

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

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

Table 1: Monte Carlo simulation results

| | $\sigma$ | 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.

of the Federal Reserve Bank of St. Louis.[10] 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 the last observation for April 2014. For brevity, we report only the results for each of the individual models with national employment data as a dependent variable.[11] 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).[12] 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

$$QPS_k^{IS} = \frac{2}{T}\sum_{t=1}^{T}(P(S_t^k = 0|\psi_t) - NBER_t)^2, \quad (34)$$

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 (i.e., $\mu_0$) 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 $\beta$. 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

[10]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.

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

[12]Note that we use equal prior probability for each model. However, we also used prior weights for each model corresponding to the fraction of employment of a given state in total national employment, which led to qualitatively similar results, suggesting that the results are robust to different specifications for the model prior probability of inclusion.

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 bivariate models, models with industrial production also tend to yield lower QPS. In particular, 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 $\alpha$ 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 jobless recoveries.

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 DMA 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 1990s, 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 of Figures 3 and 4). 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.[13] 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).

[13]The fact that BMA tends to give a weight of 1 to a single model is not very surprising. Geweke and Amisano (2011) suggest using 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 a few other models obtaining a non-negligible weight.

## Table 2: In-sample parameter estimates - Univariate models

| State | $\mu_0$ | $\mu_1$ | $\beta$ | QPS | State | $\mu_0$ | $\mu_1$ | $\beta$ | 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: $\mu_0$ 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, $\beta$ 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 (34), and the 90 per cent coverage intervals are reported in brackets.

Table 3: In-sample Parameter estimates - Bivariate models

| State | $\mu_0$ | $\mu_1$ | QPS | State | $\mu_0$ | $\mu_1$ | 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). $\mu_0$ 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 (34), and 90 per cent coverage intervals are reported in brackets.

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

| State | $\mu_0$ | $\mu_1$ | QPS | | $\mu_0$ | $\mu_1$ | 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 | Bivariate 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 bivariate models using different model-averaging schemes. $\alpha$ 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 months ahead refers to the month of March 2014). As such, our evaluation sample includes five recessions that cover 13.2 per cent of the sample. Such a long evaluation permits us 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 months 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:

$$QPS_k^{OOS} = \frac{2}{T - T_0 + 1} \sum_{t=T_0}^{T} (P(S_{t+h}^k = 0|\psi_t) - NBER_{t+h})^2, \tag{35}$$

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. The predicted probabilities of being in regime $j$ from model $k$, $P(S_{t+h}^k = j|\psi_t)$, are calculated as follows:

$$P(S_{t+h}^k = j|\psi_t) = \sum_{i=1}^{M} p_{ij}^k P(S_{t+h-1}^k = j|\psi_t), \tag{36}$$

where $M$ denotes the maximum number of regimes (two in this case) and $p_{ij}^k$ is the transition probability of going from regime $i$ to regime $j$ from model $k$ (i.e., $p_{ij}^k = P(S_{t+1}^k = j|S_t^k = i)$, calculated as the median of the parameter estimates over the 5000 simulations performed to calculate the posterior distributions of these parameters. The predicted probabilities from each model are then averaged at each point in time of the evaluation sample using the combination schemes outlined in section 3.

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. This is a very relevant benchmark, since survey forecasts have been found to perform very well compared with model-based predictions (see, e.g., Faust and Wright (2009)). The SPF is available only 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 assess 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 bivariate 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 bivariate 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 outperform 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. [14] However, an important caveat of the out-of-sample analysis so far is that we used only revised data. In the next subsection, we move to a fully real-time forecasting setting, concentrating on the prediction of the 2008-2009 recession.

### 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 available at the time the forecasts were 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. Note also that we use a real-time data series for the NBER recession dummy variable when calculating models' weights so as to carefully reflect the information available at the time the forecasts were calculated. In this purely real-time experiment, since our evaluation sample covers only 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 probability of being in a recession averaged across the different Markov-switching models (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

[14] As a robustness check, we also calculated QPS exclusively over the recession periods identified by the NBER (the results are not reported, for brevity). 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 - Bivariate 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 bivariate 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).

detail, this figure shows that the recession probability derived from the models 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 probability of recession calculated from the models 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 were very helpful to detect the beginning of the Great Recession, whereas industrial production data provided valuable information about the end of that recession.

# 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 in order 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 classification.

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 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 an important avenue for further work; however, the way seasonal adjustment should be performed remains unclear. 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


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.

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.

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

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.

Faust, J. and Wright, J. H. (2009). Comparing Greenbook and Reduced Form Forecasts Using a Large Realtime Dataset. *Journal of Business & Economic Statistics*, 27(4):468–479.

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

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

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).

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 illustrate only 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:[15]

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

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:

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

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 $\mu$, $\Sigma$, $p_{a,00}$, $p_{a,11}$, $p_{b,00}$, $p_{b,11}$, 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:

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

[15] 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 $\mu$ given $\Sigma$, $\tilde{S}_{a,T}$, $\tilde{S}_{b,T}$, 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 $\Sigma$ given $\mu$, $\tilde{S}_{a,T}$, $\tilde{S}_{b,T}$, 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

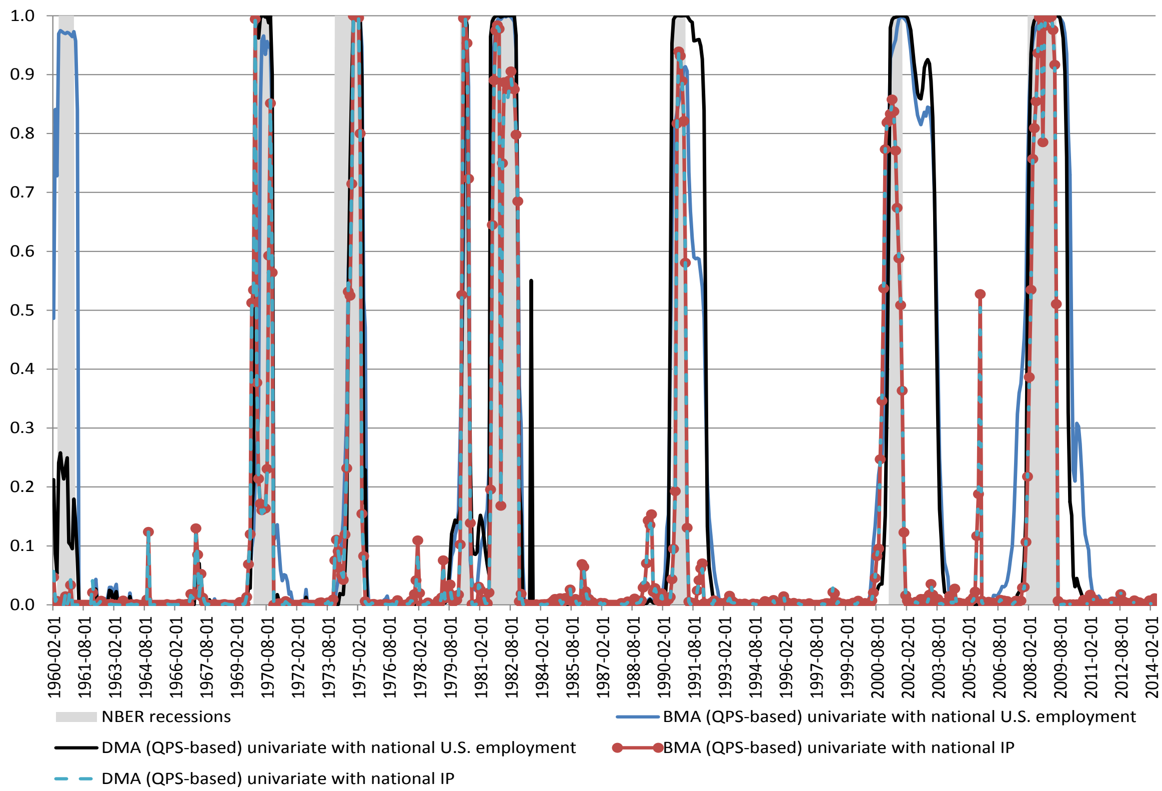
$$\Sigma^{-1}|\tilde{S}_{a,T}, \tilde{S}_{b,T}, \mu, \tilde{y}_T \sim W(\overline{S}^{-1}, \overline{v}),$$

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

After $\Sigma^{-1}$ is generated, the elements in $\Sigma$ are recovered.

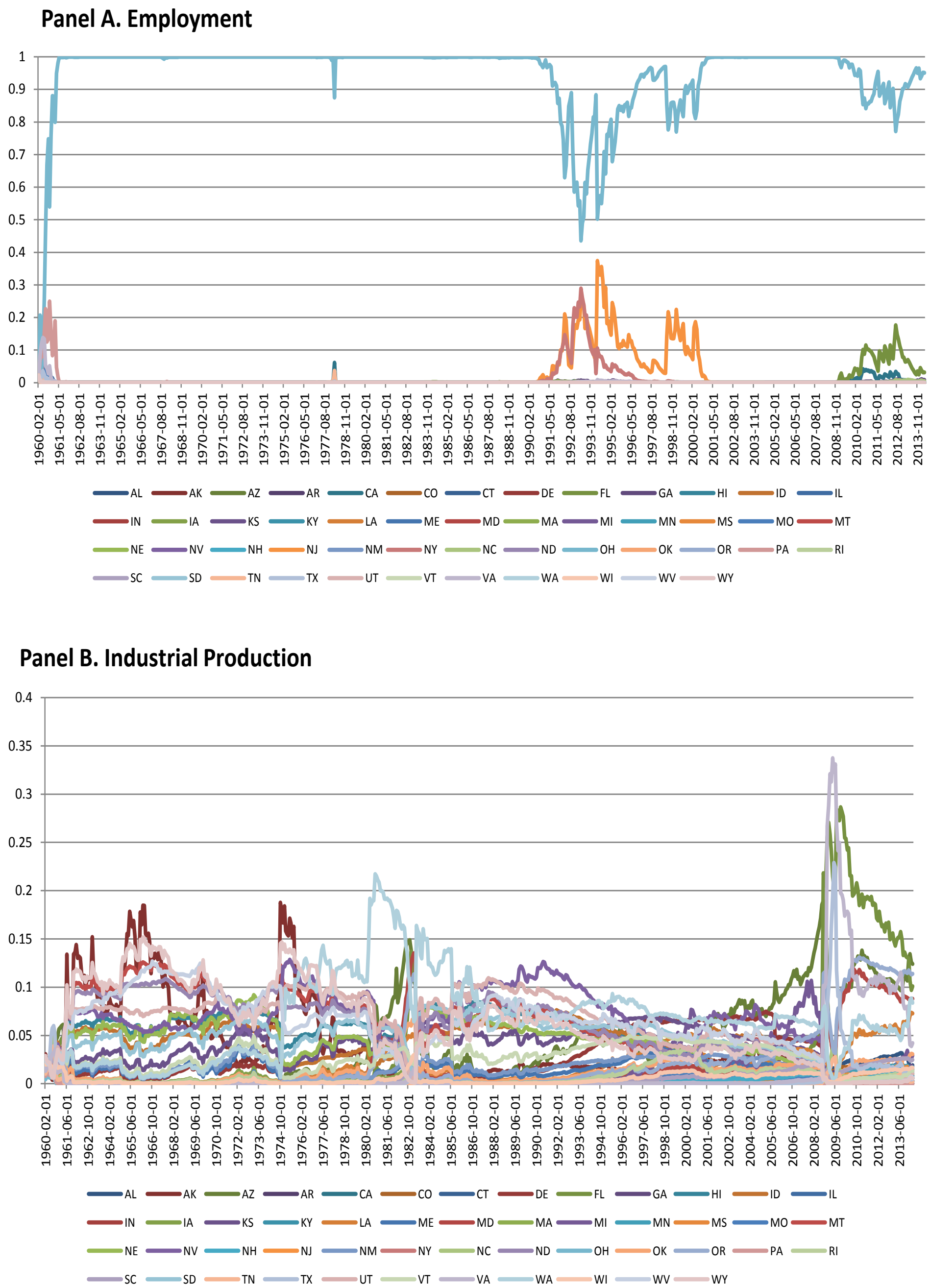
The above steps are iterated 7000 times, discarding the first 2000 iterations to mitigate the effect of the initial conditions.

Figure 1: In-sample Probability of Recession



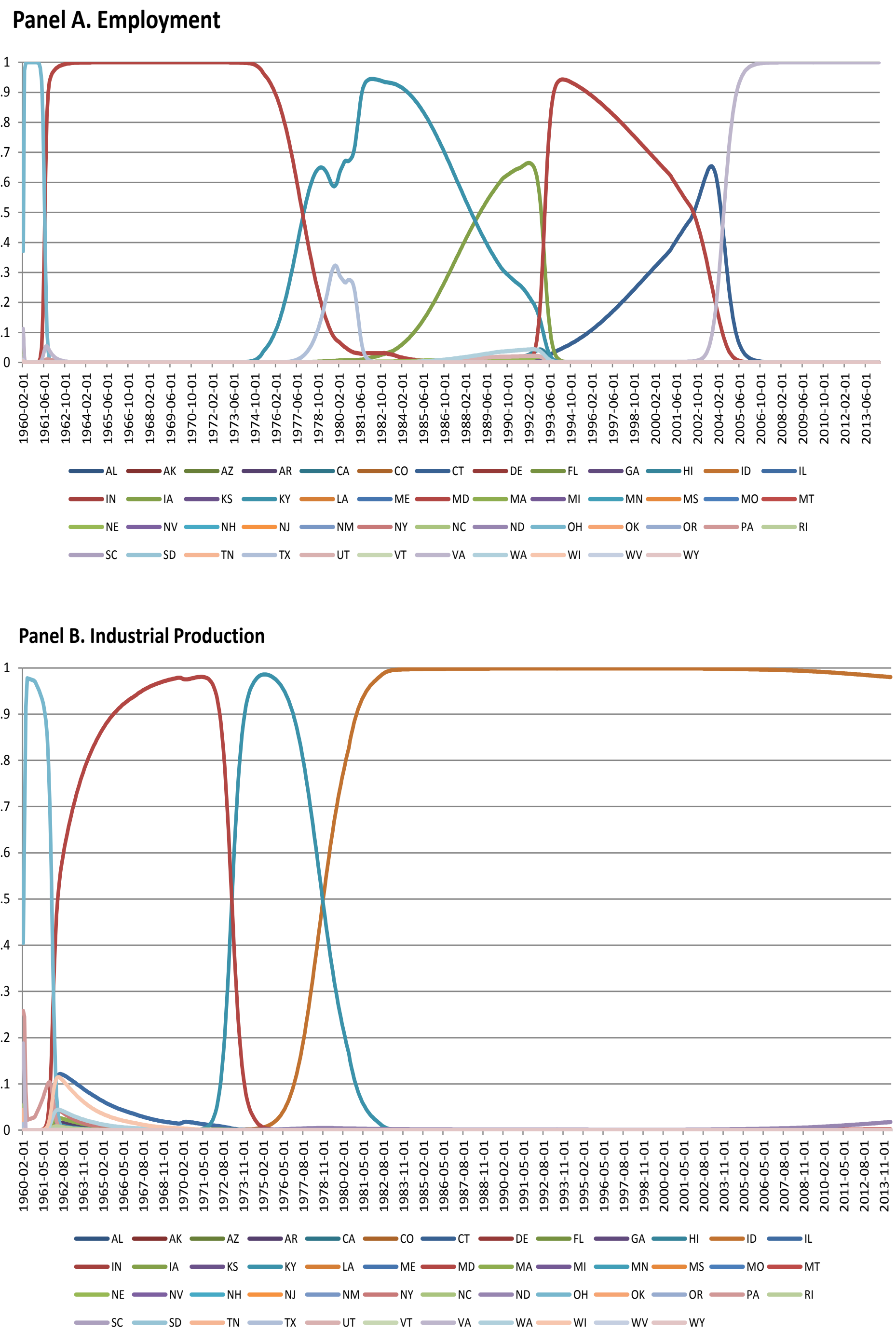

*Note:* This figure reports the monthly in-sample probability of a recession extending from February 1960 to April 2014 obtained from averaging the results from individual models using different combination schemes (BMA weights based on the QPS, and DMA weights based on the QPS).

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.

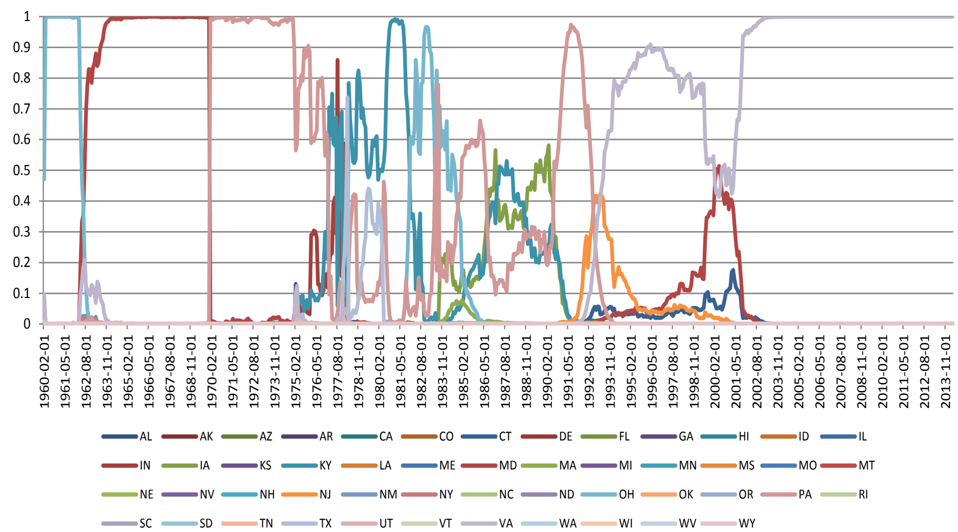
Figure 3: In-sample model weights from dynamic model averaging (QPS-based)



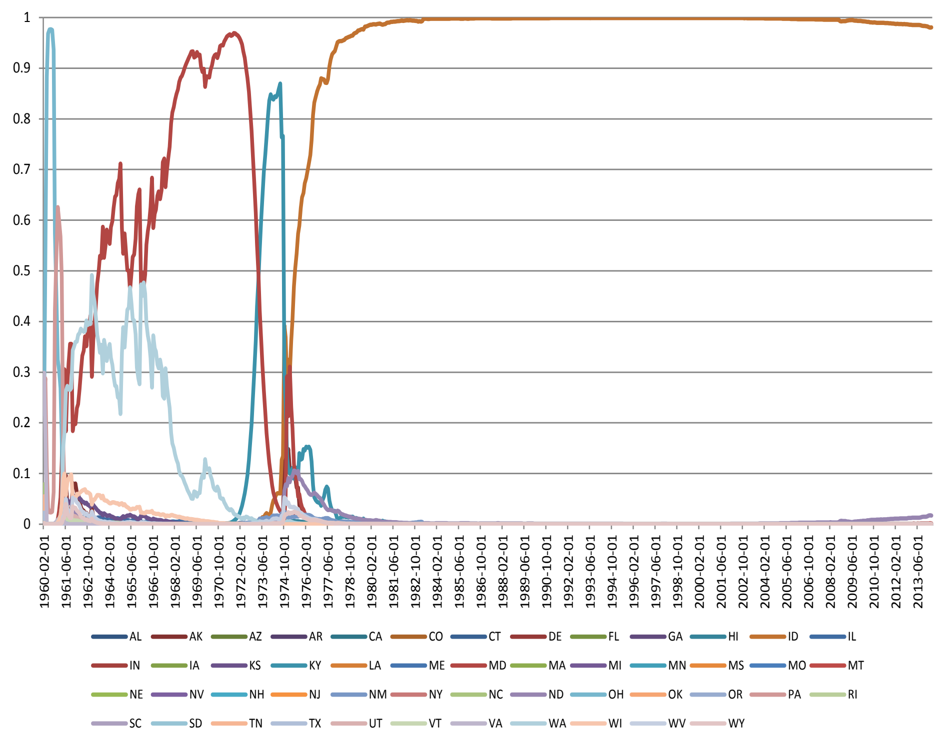

*Note:* This figure reports the weights obtained when averaging the results from univariate models using DMA weights based on the QPS (with forgetting factor $\alpha = 0.99$). Panel A shows the results when using employment as a measure of national economic activity, and panel B when using industrial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

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

**Panel A. Employment**

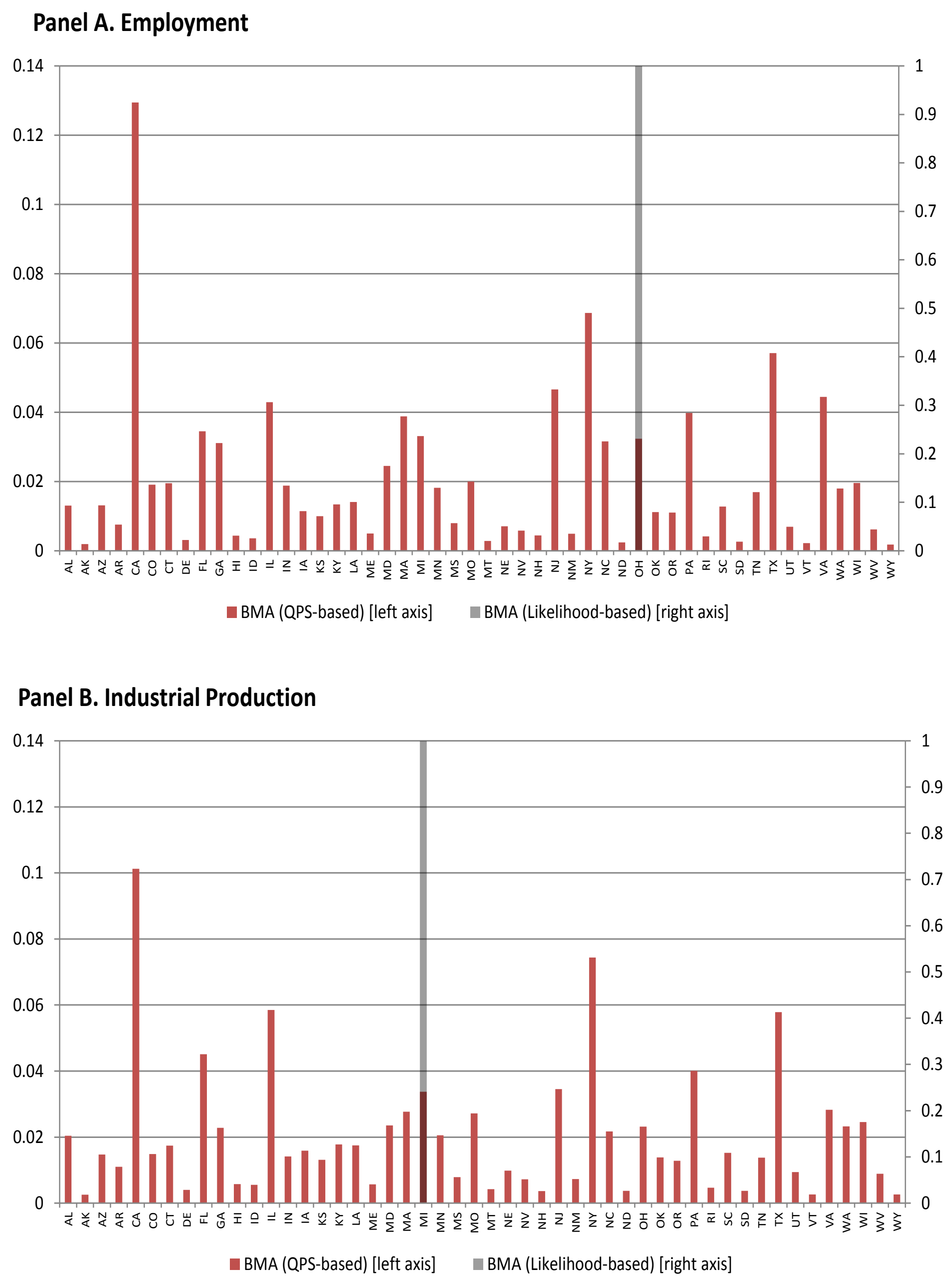



**Panel B. Industrial Production**



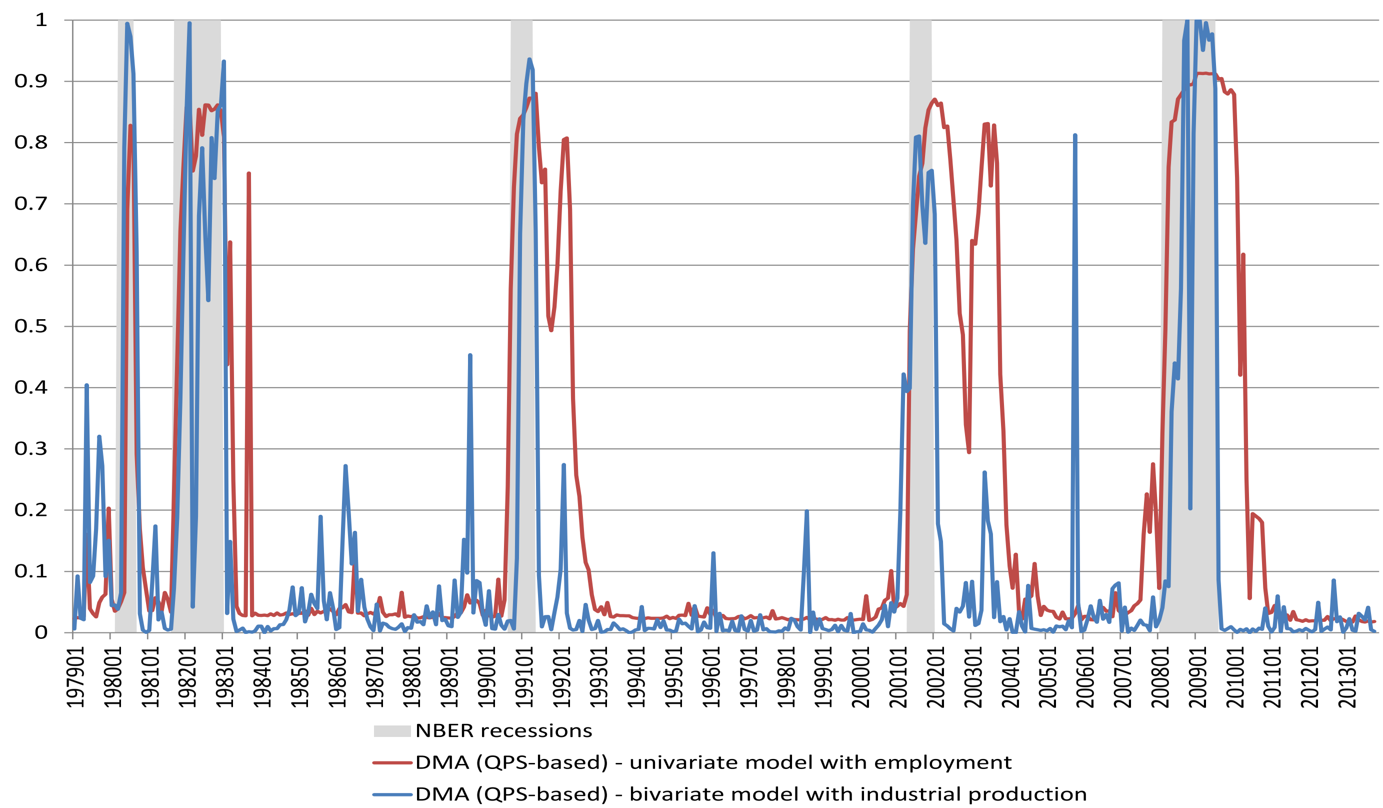

*Note:* This figure reports the weights obtained when averaging the results from univariate models using DMA weights based on a combination of QPS and predictive likelihood (with forgetting factor $\alpha = 0.99$). Panel A shows the results when using employment as a measure of national economic activity,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

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



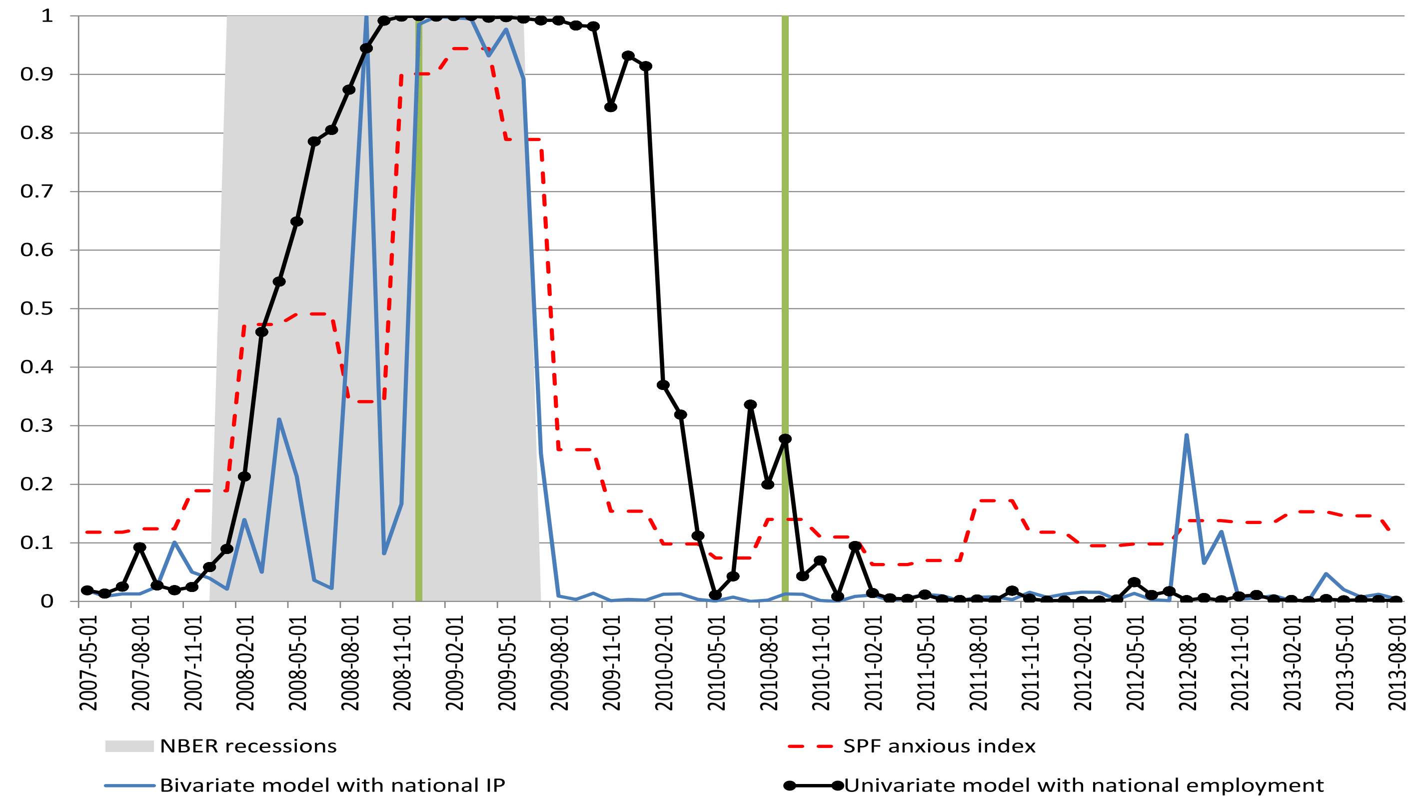

*Note:* This figure reports the weights obtained when averaging the results from univariate models using BMA weights based on the QPS and marginal likelihood. Panel A shows the results when using employment as a measure of national economic activity, and panel B when using industrial production as a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

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




*Note:* This figure reports 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 reports 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.