

# Markov-switching dynamic factor models in real time\*

Maximo Camacho

Universidad de Murcia

mcamacho@um.es

Gabriel Perez-Quiros

Banco de España and CEPR

gabriel.perez@bde.es

Pilar Poncela

Universidad Autónoma de Madrid

pilar.poncela@uam.es

November 30, 2011

## Abstract

We extend the Markov-switching dynamic factor model to account for the specificities of the day-to-day monitoring of economic developments such as ragged edges and mixed frequencies. We examine the theoretical benefits of this extension and corroborate the results by means of several Monte Carlo simulations. Finally, we assess its empirical reliability to compute real-time inferences of the US business cycle.

**Keywords:** Business Cycles, Output Growth, Time Series.

**JEL Classification:** E32, C22, E27

---

\*We are indebted to Marcelle Chauvet for graciously sharing part of the real-time data vintages used in the empirical application. Part of this paper was written while the third author was visiting the Bank of Spain. Maximo Camacho and Pilar Poncela thank the financial support of MICINN, contract grants ECO2010-19830 and ECO2009-10287, respectively. All remaining errors are our responsibility. The views in this paper are those of the authors and do not represent the views of the Bank of Spain or the Eurosystem.

# 1 Introduction

The late-2000s was the most sustained economic slump the United States has weathered since World War II. One of the lessons that this Great Recession left to economists was that policymakers and business people, who had become accustomed to the serene conditions of the Great Moderation, have dramatically incremented their interest in determining as quickly as possible whether the economy has suffered from a business cycle phase shift. In this context, time-series models, which are able to automatize the increasing complexity of the signal extraction problem in economics, facilitate the economic agents to perform and update their real-time views of the economic activity developments. These models handle with economic indicators that share the two properties of the business cycle early documented by Burns and Mitchell (1946): their signals about the economic developments are spread over the different aggregates and they exhibit business cycle asymmetries.

Diebold and Rudebusch (1996) were the first to suggest a unified model that captures these two business cycle features from a set of economic indicators. They argued that comovements among the individual economic indicators can be modelled by using the linear coincident indicator approach described in Stock and Watson (1991), while the existence of two separate business cycle regimes can be modelled by using the Markov-switching specification advocated by Hamilton (1989). Integrating these approaches, Kim and Yoo (1995), Chauvet (1998) and Kim and Nelson (1998) combined the dynamic-factor and Markov-switching frameworks to propose different versions of statistical models which capture both comovements and regime shifts at the same time. Recently, Chauvet and Hamilton (2006), Chauvet and Piger (2008), and Hamilton (2011) examine the empirical reliability of these models in computing real-time inferences of the US business cycle states.

The first important limitation of these Markov-switching dynamic factor models (MS-DFM) is that they were originally designed to deal with balanced panels of business cycle indicators. This crucial assumption leads the MS-DFM to exhibit several drawbacks when applied to the (timely) day to day monitoring of the economic activity in real time. The first drawback is related to the ragged ends that characterize the real-time data sets and that come from the typical lack of synchronicity in the daily flow of macroeconomic

information. Not accounting for this publication pattern would imply that the users of traditional MS-DFM who develop early assessments of the economic evolution from balanced panels of data will unavoidably incur into one of the following two substantial costs. The former cost appears when the forecasts are performed from the latest available balanced panel. In this case, the forecasts lose the latest and most valuable information contained in the promptly issued indicators at the time of the assessments. The later is the cost of being late when the analysts decide to wait until all the business cycle indicators become available and the inferences are then actually referred to the past.

The second limitation of the traditional MS-DFM has to do with combining information of different frequencies. This is an important limitation since some of the typical economic indicators that are observed to infer business cycle states are available at quarterly frequency while others are available at monthly frequency only. For example, the National Bureau of Economic Research (NBER) dating committee acknowledges that recessions are defined as significant declines in economic activity normally visible in real Gross Domestic Product (GDP), which is available quarterly, and real income, employment, industrial production, and wholesale-retail sales, which are available monthly. Recently, Aruoba and Diebold (2010) use *linear* dynamic factor models to account for mixed frequencies by bridging the four monthly indicators with the quarterly GDP series which is the most comprehensive measure of economic activity. However, mixing frequencies in MS-DFM is an interesting field to be explored since it typically leads to the curse of dimensionality problem. For instance, while the likelihood of a two-state DFM that only accounts for monthly indicators can be approximated by running two parallel Kalman filters in the simplest case, adding a quarterly indicator implies evaluating  $2^5 = 32$  parallel Kalman filters in the same simplest case.

The primary goal of this paper is to adapt the MS-DFM to account for these specific features that characterize the real-time monitoring of the business cycle. Our extension of these models allows economic agents that track the business cycle developments to use whatever business cycle economic indicator, regarding their publication delays and regarding their frequency. Based on the linear extension of dynamic factor models proposed by Mariano and Murasawa (2003), our procedure deals with missing observations by using

a time-varying nonlinear Kalman filter. Whenever the data is not observed, the missing observations are replaced by random draws from a variable whose distribution cannot depend on the parameter space that characterizes the Kalman filter. The corresponding row is then skipped from the Kalman recursion and the measurement equation for the missing observation is set to the random choice.

In this context, our paper provides several contributions to the existing related literature. First, we show that the incoming information provided by new releases of promptly published economic indicators is always expected to be useful to improve upon the inference about the business cycle. The expected increase in the performance accuracy is larger for less noisy indicators. Second, we point out that MS-DFM that use monthly indicators with quarterly indicators is also a good strategy to improve upon its business cycle performance. However, the accuracy gains diminish when the quality of the monthly indicators already included in the analysis is high and/or when additional quarterly indicators are noisy. In these cases, the gains from using quarterly as well as monthly indicators can be deceptively lower than expected. Third, we confirm these theoretical results by means of Monte Carlo experiments which help us to measure the magnitude of the gains from dealing with ragged ends and mixing frequencies. Fourth, we examine the ability to infer business cycle probabilities from a fast approximation algorithm that substantially reduces the curse of dimensionality problem when dealing with quarterly and monthly indicators in MS-DFM. Finally, we use a real-time data set to show that our extension of the MS-DFM leads to improvements in computing real-time business cycle inferences compared with forecasting from balanced and/or lagged panels of indicators. In particular, this is the case when interpreting the signals of business cycle phase shifts in the first months after the turning points.

The structure of this paper is organized as follows. Section 2 assesses the real-time features of the dataflow. Section 3 examines the relative performance gains of dealing with ragged ends and mixing frequencies by means of a Monte Carlo experiment. Section 4 illustrates these results for US real-time data by using the four constituent monthly series of the Stock-Watson coincident index and quarterly GDP. Section 5 concludes.

## 2 Assessing the real-time features of the data

In this section, we analyze the role of the real-time data features that characterize the real-time dataflow in the MS-DFM business cycle forecasting performance.

### 2.1 Model features

Our framework is the single-index Markov-switching dynamic factor model proposed in the mid nineties by Kim and Yoo (1995), Chauvet (1998) and Kim and Nelson (1998), that incorporates both comovements and business-cycle shifts into a statistical model. The model postulates that a vector of  $N$  economic indicators,  $\mathbf{y}_t = (y_{1,t}, \dots, y_{N,t})'$ , which are hypothesized to move contemporaneously with the overall economic conditions, can be decomposed as the sum of two components. The first component is an unobserved scalar time series variable,  $f_t$ , that accounts for the common comovements. The second component is the  $N \times 1$  time series vector  $\mathbf{u}_t$ , that represents the idiosyncratic movements in the series. This suggests the formulation:

$$\mathbf{y}_t = \mathbf{\Lambda} f_t + \mathbf{u}_t, \quad (1)$$

where  $\mathbf{\Lambda} = (\lambda_1, \lambda_2, \dots, \lambda_N)'$  is the vector of factor loadings. The main identifying assumption in the model expresses the core notion that the comovements of the multiple time series arise from the single common component. This is achieved by assuming that  $\mathbf{u}_t$  and  $f_t$  are mutually uncorrelated at all leads and lags. The remaining statistical properties  $\mathbf{u}_t$  of are stated below.

To account for the business cycle asymmetries, it is assumed that the dynamic behavior of the factor is governed by an unobserved regime-switching state variable,  $s_t$ . Within this framework, one can label  $s_t = 0$  and  $s_t = 1$  as the expansion and recession states at time  $t$ . In addition, it is standard to assume that the state variable evolves according to an irreducible 2-state Markov chain whose transition probabilities are defined by

$$p(s_t = j | s_{t-1} = i, s_{t-2} = h, \dots, I_{t-1}) = p(s_t = j | s_{t-1} = i) = p_{ij}, \quad (2)$$

where  $i, j = 0, 1$  and  $I_t$  is the information set up to period  $t$ , i.e.,  $\mathbf{y}_1, \dots, \mathbf{y}_t$ . The state variable is assumed to interact with the factor as described by the following specifications

which exhibit increasing complexity:

- (i) The common factor is a regime-switching mean

$$f_t = \mu_{s_t}. \quad (3)$$

Although it is a tight assumption that will be relaxed below,  $\mathbf{u}_t$  is assumed to be a multivariate white noise with mean equal to 0 and covariance diagonal matrix  $\Sigma_u = \text{diag}(\sigma_1^2, \sigma_2^2, \dots, \sigma_N^2)'$ . In this case, the common factor is not random when  $s_t$  and the population parameters are known. The density function of the  $k$ -th variable,  $y_{k,t}$ , does not depend explicitly on lagged values of the state variable and, conditional on the state, the covariance matrix of the observed time series does not have a factor structure. All the dynamics of the model are generated by the common regime switches.

- (ii) The common factor is governed by a regime-switching mean plus a noise

$$f_t = \mu_{s_t} + a_t, \quad (4)$$

where  $a_t$  is a white-noise process of zero mean and variance  $\sigma_a^2$ , and  $\mathbf{u}_t$  is defined as in (i). In this case, the knowledge of  $s_t$  and the population parameters still leaves remaining uncertainty about the common factor that comes from the common shocks  $a_t$ . The density function of  $y_{k,t}$  does not depend explicitly on lagged states either. Conditional on the state, the variance matrix of the observed series can be decomposed as the sum of a reduced rank matrix plus a diagonal matrix. Finally, the dynamic behavior of the model comes from the common switch.

- (iii) The common factor follows an autoregressive process with switching intercept

$$f_t = \alpha_{s_t} + \phi_1 f_{t-1} + \dots + \phi_p f_{t-p} + a_t, \quad (5)$$

and  $\mathbf{u}_t$  is defined as in (i). In this case, although the density function of  $y_{k,t}$  does not depend on lagged states, besides the dynamics implied by the common switch there is additional serial correlation given by the dependence of the common factor on its own lags. Conditional on the state, the variance and lagged covariance matrices have a factor structure.

- (iv) The common factor is given by an autoregressive process with switching mean

$$f_t = \mu_{s_t} + \phi_1 (f_{t-1} - \mu_{s_{t-1}}) + \dots + \phi_p (f_{t-p} - \mu_{s_{t-p}}) + a_t, \quad (6)$$

and  $\mathbf{u}_t$  is as in case (i). The dynamic behavior of the factor depends on the common switch and the serial dependence of the common factor on its own lags. Now, the density function of  $y_{k,t}$  depends explicitly not only on the current state but on lagged states too.

(v) The common factor is given by (4) but  $\mathbf{u}_t$  is allowed now to exhibit serial but not cross correlation. In this case, although all the common dynamics are induced by the common switch, there might be additional serial correlation which is not common across variables that is captured by the AR structure of the idiosyncratic components. This is the case that we consider in the empirical analysis.

Several additional considerations from these models deserve further comments. First, to illustrate the main theoretical results but to keep the analytical calculations as simple as possible, we assume that the order of all the autoregressive processes is always 1. Second, we skip from the analysis more complicated models that may come from using factors given by (5) or (6) where  $\mathbf{u}_t$  is allowed to exhibit serial correlation since they complicate the analytical results without adding new insights. Third, we considered models where the density function of  $y_{k,t}$  depends only on the current state and models where the density depends on lagged states as well. Finally, besides the correlation induced by the common switch, we also considered the correlation induced by autoregressive parameters that might be common or specific to each observed series.

## 2.2 Ragged ends

Ragged ends typically appear in real-time data vintages due to the unsynchronized publication of the economic indicators by the statistical agencies. Since it implies handling with unbalanced panels, the analyst who uses MS-DFM to infer the state of the business cycle in real time faces the inference problem under different alternatives that are considered in the following example. Let us assume that an analyst is trying to compute inferences about the probability of recession from the monthly indicators income, employment, industrial production, and sales on the 15th day of month  $t + 2$ . Recall that income and sales exhibit publication delays of one and two months, respectively.

The first alternative consists on computing the business cycle inferences from the balanced panels. In this context, the either the inferred probabilities at  $t + 2$  are actually

referring to the past,  $prob(s_t = 1|I_t)$ , or the probabilities are actually obtained with a delay of two months since they can only be computed at  $t + 4$ ,  $prob(s_{t+2} = 1|I_{t+4})$ , when the statistical agencies publish the latest releases for those indicators that exhibit publication lags and the panels become balanced. However, neither of these strategies is very useful in practice. In the first case, the economic agents are forced to adopt their decisions at  $t + 2$  as if the economic conditions were identical to those occurred at  $t$ . In the second case, business people are exposed to suffer from the dangers of business cycle phase shifts occurred during the waiting times  $t + 2$  and  $t + 3$ . Accordingly, this alternative is not considered in the paper.

The second alternative consists on computing forecasts for  $t+2$  from the latest available balanced panel at  $t$ . Since the probabilities quickly collapse to their ergodic values, it typically implies that one can badly infer the current state in real time. In our example, computing recession probabilities on the 15th day of month  $t + 2$  would require two-period-ahead forecasts from the complete panel which contains information up to  $t$ , i.e.,  $prob(s_{t+2} = 1|I_t)$ . Let us assume that  $p_{00} = 0.9$  and  $p_{11} = 0.7$ , which are the percentage of quarters classified as expansions that are followed by expansions and the percentage of quarters classified as recessions that are followed by recessions in the period 1959.3-2010.3 by the NBER, respectively. Let us also assume that  $prob(s_t = 1|I_t) = 1$ , i.e., the economy is surely in recession at  $t$ . In this case, the recession probability inferred at  $t + 2$  using the information available up to  $t$ , becomes

$$prob(s_{t+2} = 1|I_t) = \sum_{i=0}^1 p_{1i}p_{i1}prob(s_t = 1|I_t) \quad (7)$$

or a coin flip probability of 0.52 only. Hence, the probability generated at  $t + 2$  with information up to  $t$  is only consistent with a non definitive suspicion that a recession remains, although no additional information has been incorporated either in  $t + 1$  or in  $t + 2$ .

To understand the benefits of our method, it is worth recalling that in the previous example sales is only available up to  $t$  on the 15th day of month  $t$ , income is available up to  $t + 1$ , and industrial production and employment are available up to  $t + 2$ . Hence, informative news about turning points occurring from  $t$  to  $t+2$  such as dramatic changes in

stock markets, in economic policies, in the financial system or wars and natural disasters, would be reflected in the incoming publication of the business cycle indicators and would help the economic agents to form clearer signals about the current state of the business cycle than the probability predicted with data up to  $t$ . According to this example, using at  $t + 2$  the marginal information published at  $t + 1$  and  $t + 2$  that help to improve upon the inferred probability  $prob(s_{t+2} = 1|I_t)$  up to  $prob(s_{t+2} = 1|I_{t+2})$  would require the MS-DFM to deal with unbalanced panels.

The previous example describes a very realistic empirical situation since, regarding the economic region considered in the business cycle analysis, some frequently used hard data usually exhibit publication lags of up to two months. Although the two-month publication delay will be treated carefully in the simulations and in the empirical analysis, to facilitate the theoretical analysis we consider publication lags of one month only.<sup>1</sup>

In the analysis of how the upcoming information can help the models to compute business cycle inferences, we distinguish two cases: models with switching intercepts and models with switching means. Both have been used in the related literature. For instance, the MS-DFM with switching intercepts has been used by Kim and Yoo (1995), Chauvet (1998) and Chauvet and Hamilton (2006), while the MS-DFM with switching means has been used by Kim and Nelson (1998) and Chauvet and Piger (2008).

### **2.2.1 First arrival at $t + 1$ in a model with switching intercept**

In a broad sense, we consider a model as a MS-DFM with switching intercepts if the probabilistic density function of each indicator  $y_{k,t}$  depends explicitly on the current state,  $s_t$ , but not on its lags. It simplifies the estimation of the nonlinear model and the inference about the state probabilities. The cases (i) to (iii) belong to this type of MS-DFM.

Let us assume that the analyst is restricted to use balanced panels as in traditional MS-DFM. Using the information  $I_t$  obtained before the first arrival of new information at  $t + 1$ , her best guess of the probability of being in a certain state at  $t + 1$  is the one step

---

<sup>1</sup>The extension to larger lags is conceptually easy.

ahead forecast of that probability

$$prob(s_{t+1} = j|I_t) = \sum_{i=0}^1 prob(s_t = i|I_t)p_{ij}. \quad (8)$$

For example, for the numbers of the example described above, these state probabilities are  $prob(s_{t+1} = 0|I_t) = 0.3$  and  $prob(s_{t+1} = 1|I_t) = 0.7$ .

However, let us assume now that the analyst uses the generalization of MS-DFM proposed in this paper which incorporates the information provided by some promptly published business cycle indicator,  $y_{k,t+1}$ , that exhibits no publication lags. In this case, the analyst can incorporate this information as it arrives and update her belief of being in a certain state at  $t + 1$  as

$$prob(s_{t+1} = j|I_t, y_{k,t+1}) = \frac{f(y_{k,t+1}|s_{t+1} = j, I_t)}{f(y_{k,t+1}|I_t)}prob(s_{t+1} = j|I_t). \quad (9)$$

Using the new information helps the analysts if it rises the ability to increase the true signals. This implies that it should rise the probability of a given state when the economy is actually in that state. For instance, let us assume that the economy is in recession at  $t + 1$ , i.e.,  $s_{t+1} = 1$ ,<sup>2</sup> and that  $0 < prob(s_{t+1} = j|I_t) < 1$ .<sup>3</sup> Hence, the partial information provided by  $y_{k,t+1}$  is helpful to reduce false signals when  $prob(s_{t+1} = 1|I_t, y_{k,t+1}) > prob(s_{t+1} = 1|I_t)$  if  $s_{t+1} = 1$ , which occurs whenever

$$f(y_{k,t+1}|s_{t+1} = 1, I_t) > f(y_{k,t+1}|I_t). \quad (10)$$

Using the total law of probabilities, if  $s_{t+1} = 1$  the condition in (10) would be

$$f(y_{k,t+1}|s_{t+1} = 1, I_t) > f(y_{k,t+1}|s_{t+1} = 1, I_t)prob(s_{t+1} = 1|I_t) + f(y_{k,t+1}|s_{t+1} = 0, I_t)prob(s_{t+1} = 0|I_t), \quad (11)$$

which, rearranging terms, leads to

$$f(y_{k,t+1}|s_{t+1} = 1, I_t) > f(y_{k,t+1}|s_{t+1} = 0, I_t). \quad (12)$$

As Hamilton (2011) illustrates for univariate processes, the condition about the probabilities can be viewed as a condition about the height of the conditional densities of the observed variable.

---

<sup>2</sup>Of course, we obtain the analogous results for the probability of expansion if we assume that the true state is  $s_{t+1} = 0$ .

<sup>3</sup>It is straightforward to check that if  $prob(s_{t+1} = j|I_t) = 1$ , then  $prob(s_{t+1} = j|I_t, y_{k,t+1}) = 1, j = 0, 1$ .

To easily interpret this condition in the particular cases of switching intercepts, it is easy to show that for conditional Gaussian densities with equal variance in both states and assuming that  $\mu_1 < \mu_0$  (i.e., the average growth rate is lower in recessions than in expansions), the condition is satisfied if

$$(y_{k,t+1} - \lambda_k \mu_1)^2 < (y_{k,t+1} - \lambda_k \mu_0)^2 \quad (13)$$

for cases (i) and (ii), and if  $\alpha_1 < \alpha_0$ , it is satisfied if

$$(y_{k,t+1} - \lambda_k f_{t+1|t}^{(1)})^2 < (y_{k,t+1} - \lambda_k f_{t+1|t}^{(0)})^2 \quad (14)$$

for case (iii). In the last expression,  $f_{t+1|t}^{(i)} = \alpha_i + \phi f_{t|t}$  is the conditional expectation of the common factor given  $s_{t+1} = i$  and the observed series up to time  $t$ , where  $i = 0, 1$ , and  $f_{t|t} = E(f_t | I_t)$  is the conditional expectation of the common factor given the observed series up to time  $t$ . These conditions tell the analyst that, to decide whether a particular value of  $y_{k,t+1}$  is helpful to know the probability of being in state 1 in  $s_{t+1}$  when the economy actually is in that state, it helps to compute the squared forecast errors of  $y_{k,t+1}$  associated to each of the states and to check if the squared forecast error associated to  $s_{t+1} = 1$  is smaller than the squared forecast error associated to  $s_{t+1} = 0$ .

In addition, the conditions can also be interpreted by using the classification rules obtained in linear discriminant analysis with constant variances in both groups. After some straightforward algebra in the last two expressions, it is easy to show that if  $s_{t+1} = 1$ , then using  $y_{k,t+1}$  is helpful to reduce false signals when

$$y_{k,t+1} < \lambda_k \frac{(\mu_0 + \mu_1)}{2} \quad (15)$$

for cases (i) and (ii), and

$$y_{k,t+1} < \lambda_k \left( \frac{\alpha_0 + \alpha_1}{2} + \phi f_{t|t} \right) \quad (16)$$

for case (iii). In this case, incorporating the new information  $y_{k,t+1}$  is helpful to infer the state  $s_{t+1} = 1$  if the new observation is lower than the average between the means in the two states.

In practice, the two density functions that appear in (12) may overlap and the condition does not hold for all values of  $y_{k,t+1}$ . In these cases, even when the true state is  $s_{t+1} = 1$ ,

it might happen that for some (usually high) values of  $y_{k,t+1}$ ,  $f(y_{k,t+1}|s_{t+1} = 1, I_t) < f(y_{k,t+1}|s_{t+1} = 0, I_t)$ . This might lead to a false signal detection since the probability for  $s_{t+1} = 1$  decreases when  $y_{k,t+1}$  is observed. Accordingly, the usefulness of  $y_{k,t+1}$  to compute business cycle inferences must be evaluated on average. Taking natural logarithms, the condition in (12) implies that using the incoming information helps in inferring recession probabilities in actual recessions when

$$\ln f(y_{k,t+1}|s_{t+1} = 1, I_t) - \ln f(y_{k,t+1}|s_{t+1} = 0, I_t) > 0. \quad (17)$$

Now, taking into account all possible outcomes of  $y_{k,t+1}$  for the true state  $s_{t+1} = 1$ , the expected value of the difference between the two conditional densities under conditional Gaussianity is given by the Kullback-Leibler divergence

$$\int \ln \frac{f(y_{k,t}|s_{t+1} = 1, I_t)}{f(y_{k,t}|s_{t+1} = 0, I_t)} f(y_{k,t+1}|s_t = 1, I_t) dy_{k,t+1}. \quad (18)$$

The next proposition, which is based on the concept of conditional entropy, quantifies the potential advantage of adding the advanced business cycle signal provided by the promptly published indicator  $y_{k,t+1}$ .

**Proposition 1** *Assume the factor model given by (1) for the observed series and the common factor considered in cases (i) to (iii). Under conditional Gaussianity, the gains in  $t + 1$  from using the observed  $y_{k,t+1}$  to infer the business cycle, given by the Kullback-Leibler (KL) divergence of  $f(y_{k,t+1}|s_{t+1} = 1, I_t)$  with respect to  $f(y_{k,t+1}|s_{t+1} = 0, I_t)$  are*

$$KL(i) = \frac{\lambda_k^2(\mu_0 - \mu_1)^2}{2\sigma_k^2}, \quad (19)$$

$$KL(ii) = \frac{\lambda_k^2(\mu_0 - \mu_1)^2}{2(\sigma_a^2\lambda_k^2 + \sigma_k^2)}, \quad (20)$$

$$KL(iii) = \frac{\lambda_k^2(\alpha_0 - \alpha_1)^2}{2\sigma_{k,t+1|t}^2}, \quad (21)$$

where  $\sigma_{k,t+1|t}^2$  is the conditional one step ahead error variance of  $y_{k,t+1}$ , which has been assumed to be the same in both states.

**Proof:** The proof is given in the Appendix.

These expressions imply that if there are separate business cycles regimes in the sense that  $\mu_1 \neq \mu_0$ ,  $\alpha_0 \neq \alpha_1$ , and  $\sigma_a^2 < \infty$ , and if the new information provided by the early published indicator is informative in the sense that  $\lambda_k \neq 0$ ,  $\sigma_k^2 < \infty$ , and  $\sigma_{k,t+1|t}^2 < \infty$ , then the divergence is strictly positive. This implies that the incoming information provided by an informative new release of the  $k$ -th economic indicator is always expected to be useful to improve upon the inference about the business cycle at time  $t + 1$ .

It is worth emphasizing that although the Gibbs inequality guarantees that the difference in the conditional densities is always positive, this proposition quantifies the magnitude of the change in the conditional entropy and measures the averaged business cycle information content of  $y_{k,t+1}$ . Basically, the information content of  $y_{k,t+1}$  increases with the difference between the within-state means and with the ability of the new issue to capture this information in the model, which is given by the conditional signal-to-noise ratio  $\frac{\lambda_k^2}{\sigma_{k,t+1|t}^2}$ . In particular, the proposition shows that the greater the signal  $\lambda_k^2$ , and the lower the one step ahead noise variance  $\sigma_{k,t+1|t}^2$ , the larger the expected gains from observing  $y_{k,t+1}$ .

Regarding the noise of the signals, the three models consider state-independent variance of the observed indicator, i.e.,  $(\sigma_k^{(j)})^2 = \sigma_k^2$ ,  $j = 0, 1$ . This also implies that the conditional one step ahead forecast error variance of  $y_{k,t+1}$  is state-independent,  $(\sigma_{k,t+1|t}^{(j)})^2 = \sigma_{k,t+1|t}^2$ ,  $j = 0, 1$ . In this context, it is worth emphasizing that  $\sigma_{k,t+1|t}^2$  does not denote here the variance of the Gaussian mixture at  $t + 1|t$  but the variance of either of the components of the mixture at  $t + 1|t$ , which is given by

$$\sigma_{k,t+1|t}^2 = \lambda_k^2 (\phi^2 P_{t|t} + \sigma_a^2) + \sigma_k^2, \quad (22)$$

where  $P_{t|t} = E(f_t - f_{t|t}|I_t)^2$ . However, the results can be easily generalized to the case of different variances across states.

The previous results can also be extended to the case where several business cycle indicators are published with very short delay and therefore are available at  $t + 1$ . For this purpose, let  $\mathbf{y}_{k,t+1}$  be the subvector of available indicators at  $t + 1$ , and let  $\Sigma_{t+1|t}$  be the variance-covariance matrix of the one step ahead forecast errors of this subvector. In addition, let  $\mathbf{\Lambda}_k$  be the subvector of factor loadings that corresponds to the series in

$\mathbf{y}_{k,t+1}$ , and let  $\phi$  be the autoregressive parameter that drives the common factor dynamics. Under Gaussianity of the conditional densities, incorporating in the model the incoming information of the subvector  $\mathbf{y}_{k,t+1}$  will increase the accuracy to detect a recession when

$$\begin{aligned} & (\mathbf{y}_{k,t+1} - \mathbf{\Lambda}_k(\alpha_0 + \phi f_{t|t}))' \boldsymbol{\Sigma}_{t+1|t}^{-1} (\mathbf{y}_{k,t+1} - \mathbf{\Lambda}_k(\alpha_0 + \phi f_{t|t})) \\ & > (\mathbf{y}_{k,t+1} - \mathbf{\Lambda}_k(\alpha_1 + \phi f_{t|t}))' \boldsymbol{\Sigma}_{t+1|t}^{-1} (\mathbf{y}_{k,t+1} - \mathbf{\Lambda}_k(\alpha_1 + \phi f_{t|t})). \end{aligned} \quad (23)$$

Hence, the new data incorporate informational content if the Mahalanobis distance between the new (corrected of serial dynamics) data to the conditional expectation in state 0 is greater than the distance of the data to the conditional expectation in state 1.

After some straightforward algebra, it can be shown that the condition is achieved whenever

$$\boldsymbol{\Lambda}'_k \boldsymbol{\Sigma}_{t+1|t}^{-1} \mathbf{y}_{k,t+1} < \frac{\alpha_1 + \alpha_0}{2} \boldsymbol{\Lambda}'_k \boldsymbol{\Sigma}_{t+1|t}^{-1} \mathbf{\Lambda}_k. \quad (24)$$

Using the notation of linear discriminant analysis, the condition holds when the scalar variable  $x_{t+1} = \boldsymbol{\beta}' \mathbf{y}_{k,t+1}$  is closer to  $m_1 = \boldsymbol{\beta}' \mathbf{\Lambda}_k \alpha_1$  than to  $m_0 = \boldsymbol{\beta}' \mathbf{\Lambda}_k \alpha_0$ , where  $\boldsymbol{\beta} = \boldsymbol{\Sigma}_{t+1|t}^{-1} \mathbf{\Lambda}_k (\alpha_1 - \alpha_0)$ . If the new information is very volatile, meaning that  $\boldsymbol{\Sigma}_{t+1|t}$  is large or the two intercepts are not far enough from each other, it may happen that the new information increases the misclassification rate.

To evaluate (21) for all possible outcomes of the subvector  $\mathbf{y}_{k,t+1}$  when the true state is  $s_{t+1} = 1$ , we proceed as in the scalar case and compute the corresponding Kullback-Leibler divergence. The results are given in the following proposition.

**Proposition 2** *Assume the factor model given by (1) with the common factor given as in (iii). Under conditional Gaussianity, the gain in  $t + 1$  from observing the subvector of  $k$  variables  $\mathbf{y}_{k,t+1}$ , given by the Kullback-Leibler divergence of  $f(\mathbf{y}_{k,t+1}|s_{t+1} = 1, I_t)$  with respect to  $f(\mathbf{y}_{k,t+1}|s_{t+1} = 0, I_t)$ , is given by*

$$KL = \frac{(\alpha_0 - \alpha_1)^2}{2} \boldsymbol{\Lambda}'_k \boldsymbol{\Sigma}_{t+1|t}^{-1} \mathbf{\Lambda}_k, \quad (25)$$

where  $\boldsymbol{\Sigma}_{t+1|t}$  is the conditional variance-covariance matrix of the one step ahead forecast error of  $\mathbf{y}_{k,t+1}$ , which is the same in both states, and  $\mathbf{\Lambda}_k$  is the subvector of factor loadings that corresponds to the series in  $\mathbf{y}_{k,t+1}$ .

**Proof:** The proof is given in the Appendix.

Notice that the expressions for the other two cases can be easily obtained from case (iii) by imposing  $\phi = 0$ , and  $\mu_i = \alpha_i$ ,  $i = 0, 1$  in case (ii), and by imposing the additional restriction  $\text{var}(a_t)=0$  in case (i).

### 2.2.2 First arrival at $t + 1$ in models with switching mean

Now, instead of allowing the intercept term of the common component to be regime dependent, we allow the mean of the factor to be regime dependent. In this case,  $\mathbf{y}_{t+1}$  depends not only on  $s_{t+1}$  but also on the lagged business cycle states. The case (iv) belongs to this type of MS-DFM.

To keep the analysis as simple as possible, we assume that  $\mathbf{y}_t$  is generated by (1), where  $\mathbf{u}_t$  is assumed to be a multivariate white noise with mean equal to 0 and covariance diagonal matrix  $\Sigma_u = \text{diag}(\sigma_1^2, \sigma_2^2, \dots, \sigma_N^2)'$ , and  $f_t$  is given by (6). The model can be expressed as

$$\mathbf{y}_t = \mathbf{\Lambda}\mu_{s_t} + \phi\mathbf{\Lambda}(f_{t-1} - \mu_{s_{t-1}}) + \mathbf{\Lambda}a_t + \mathbf{u}_t. \quad (26)$$

It is worth noting that the density function of  $\mathbf{y}_t$  depends not only on  $s_t$  but also on its first lag  $s_{t-1}$ .

The forecasted probability of being in state  $j$  at  $t + 1$  computed from a MS-DFM which is restricted to use a balanced panel of indicators with information  $I_t$ , is computed from

$$\text{prob}(s_{t+1} = j|I_t) = \sum_{i=0}^1 \text{prob}(s_{t+1} = j, s_t = i|I_t). \quad (27)$$

However, if the information of an early published economic indicator  $y_{k,t+1}$  is incorporated into the model by using the generalization of the MS-DFM proposed in this paper, the inference about the state probabilities is computed from

$$\text{prob}(s_{t+1} = j|I_t, y_{k,t+1}) = \sum_{i=0}^1 \frac{f(y_{k,t+1}|s_{t+1} = j, s_t = i, I_t)}{f(y_{k,t+1}|I_t)} \text{prob}(s_{t+1} = j, s_t = i|I_t), \quad (28)$$

The difference between (27) and (28) is given by the ratio of conditional densities that premultiply the joint probabilities in (28) and that equals one in (27).

For a particular point evaluation of the performance of  $y_{k,t+1}$  in separating the two states, under the assumption of conditional Gaussianity and state-independent one-step-ahead error variances, a sufficient condition for  $\text{prob}(s_{t+1} = j|I_t, y_{k,t+1}) > \text{prob}(s_{t+1} = j|I_t)$  when  $s_{t+1} = j$  is given by

$$f(y_{k,t+1}|s_{t+1} = j, s_t = i, I_t) > f(y_{k,t+1}|I_t), \quad (29)$$

for  $i = 0, 1$ . Hence,  $y_{k,t+1}$  exhibits informational content when including its new information in the model if it tends to favor the particular path  $(i, j)$ , being  $i$  the state at  $t$  and  $j$  the state at  $t + 1$ . In particular, we say that  $y_{k,t+1}$  favors the path  $(i, j)$  if under conditional Gaussianity

$$(y_{k,t+1} - \lambda_k \mu_j - \phi \lambda_k (f_{t|t}^{(i)} - \mu_i))^2 < (y_{k,t+1} - \lambda_k \mu_m - \phi \lambda_k (f_{t|t}^{(l)} - \mu_l))^2 \quad (30)$$

for all pairs  $(l, m) \neq (i, j)$ ,  $i, j, k, l = 0, 1$ , where  $f_{t|t}^{(n)} = E(f_t|s_t = n, I_t)$ . As in the case of switching intercepts, the condition implies that  $y_{k,t+1}$  exhibits informational content when it increases the forecasting performance of the model within the true state.

The condition can also be interpreted by using classification rules. After some straightforward algebra, the condition ensuring that  $y_{k,t+1}$  has informational content to compute business cycle inferences can be stated as

$$y_{k,t+1} < \lambda_k \frac{\mu_j + \phi(f_{t|t}^{(i)} - \mu_i) + \mu_m + \phi(f_{t|t}^{(l)} - \mu_l)}{2}, \quad (31)$$

which takes into account the direct dependence of  $y_{k,t+1}$  on  $s_{t+1}$  and the state from which the economy arrives.

This result can also be extended to the case of  $\mathbf{y}_{k,t+1}$  being a subvector of  $\mathbf{y}_{t+1}$ , instead of one single time series only. Under Gaussianity of the conditional densities and state-independent variances, the subvector of promptly published economic indicators  $\mathbf{y}_{k,t+1}$  exhibits informational content to infer the states  $s_t = i$  and  $s_{t+1} = j$  if

$$\begin{aligned} & (\mathbf{y}_{k,t+1} - \mathbf{\Lambda}_k \mu_j - \phi \mathbf{\Lambda}_k (f_{t|t}^{(i)} - \mu_i))' \mathbf{\Sigma}_{k,t+1|t}^{-1} (\mathbf{y}_{k,t+1} - \mathbf{\Lambda}_k \mu_j - \phi \mathbf{\Lambda}_k (f_{t|t}^{(i)} - \mu_i)) \\ & < (\mathbf{y}_{k,t+1} - \mathbf{\Lambda}_k \mu_m - \phi \mathbf{\Lambda}_k (f_{t|t}^{(l)} - \mu_l))' \mathbf{\Sigma}_{k,t+1|t}^{-1} (\mathbf{y}_{k,t+1} - \mathbf{\Lambda}_k \mu_m - \phi \mathbf{\Lambda}_k (f_{t|t}^{(l)} - \mu_l)). \end{aligned} \quad (32)$$

Hence, it is worth using the generalized MS-DFM that accounts for ragged ends when the Mahalanobis distance between the new data, corrected of idiosyncratic dynamics, for  $(s_t, s_{t+1}) = (i, j)$  is lower than for any other pair  $(s_t, s_{t+1}) \neq (i, j)$ .

The overall informational content of using promptly published indicators can also be computed by integrating over all its possible outcomes. As in the case of common factors with switching intercepts, the averaged better performance can be measured by taking natural logarithms of the conditional densities ratio and integrating according to the true distribution that appears in the true path,  $s_t = i$ ,  $s_{t+1} = j$ . The Kullback-Leibler divergence becomes in this case

$$\int \ln \frac{f(y_{k,t+1}|s_{t+1} = j, s_t = i, I_t)}{f(y_{k,t+1}|I_t)} f(y_{k,t+1}|s_{t+1} = j, s_t = i, I_t) dy_{k,t+1}. \quad (33)$$

However, a new problem appears in this case since it implies evaluating the integral when the distribution is a mixture. In the spirit of Harrison and Stevens (1976) and Peña and Guttman (1989) we use the ‘‘collapsing’’ method that approximates a mixture of Gaussians by a Gaussian with the same mean and variance of the mixture.<sup>4</sup>

In the context of entropy calculations, this problem has been treated in Hershey and Olsen (2007). We benefited from their method to approximate the entropy of a mixture of Gaussians to deal with the following proposition.

**Proposition 3** *Assume the factor model given by (1) for the observed series and the common factor considered in case (iv). Under conditional Gaussianity, the gain in  $t + 1$  from observing  $y_{k,t+1}$ , given by the Kullback-Leibler divergence of  $f(y_{k,t+1}|s_{t+1} = j, s_t = i, I_t)$  with respect to  $f(y_{k,t+1}|I_t)$  can be approximated by*

$$KL(iv) = \frac{1}{2} \ln \frac{\tilde{\sigma}_{k,t+1|t}^2}{\sigma_{k,t+1|t}^2} \quad (34)$$

$$= \frac{1}{2} \ln \left( 1 + \sum_{i=0}^1 \sum_{j=0}^1 \text{prob}(s_{t+1} = j, s_t = i|I_t) (y_{k,t+1|t}^{(i,j)} - \tilde{y}_{k,t+1|t})^2 \right), \quad (35)$$

where  $y_{k,t+1|t}^{(i,j)} = E(y_{k,t+1}|s_{t+1} = j, s_t = i, I_t)$ ,  $i, j = 0, 1$ , is given by

$$y_{k,t+1|t}^{(i,j)} = \lambda_k \mu_j - \phi \lambda_k (f_{t|t}^{(i)} - \mu_i), \quad (36)$$

with  $f_{t|t}^{(i)} = E(f_t|s_t = i, I_t)$ , and  $\tilde{y}_{k,t+1|t}$  and  $\tilde{\sigma}_{k,t+1|t}^2$  are the mean and variance of the

---

<sup>4</sup>The collapsing approach has also been used in the context of Markov-switching algorithms by Kim (1994), among others.

mixture given by

$$\tilde{y}_{k,t+1|t} = \sum_{i=0}^1 \sum_{j=0}^1 \text{prob}(s_{t+1} = j, s_t = i | I_t) y_{k,t+1|t}^{(i,j)}, \quad (37)$$

$$\tilde{\sigma}_{k,t+1|t}^2 = \sigma_{k,t+1|t}^2 + \sum_{i=0}^1 \sum_{j=0}^1 \text{prob}(s_{t+1} = j, s_t = i | I_t) (y_{k,t+1|t}^{(i,j)} - \tilde{y}_{k,t+1|t})^2. \quad (38)$$

**Proof:** The proof is given in the Appendix.

Proposition 3 quantifies the prediction or information content of  $y_{k,t+1}$  for the type of models considered in (iv).<sup>5</sup> It depends on the dispersion of the observed data and the probability associated to each possible path.

Finally, we consider case (v) where the common factor is given as in (4), but each idiosyncratic component is an autoregressive process of order one,

$$\mathbf{u}_t = \Psi \mathbf{u}_t + \boldsymbol{\epsilon}_t, \quad (39)$$

where  $\text{var}(\boldsymbol{\epsilon}_t) = \text{diag}(\sigma_1^2, \dots, \sigma_N^2)$  and  $\Psi = \text{diag}(\psi_1, \dots, \psi_N)$ . In this case, a particular path  $(i, j)$  is favored on average when the true states is precisely  $(i, j)$  if a similar condition to (33) holds. The next proposition quantifies the prediction or information content of  $y_{k,t+1}$ .

**Proposition 4** *Assume the factor model given by (1) for the observed series and the common factor considered in case (v). Under conditional Gaussianity, the gain in  $t+1$  from observing  $y_{k,t+1}$ , given by the Kullback-Leibler divergence of  $f(y_{k,t+1} | s_{t+1} = i, s_t = j, I_t)$  with respect to  $f(y_{k,t+1} | I_t)$ , can be approximated by*

$$KL(v) = \frac{1}{2} \ln \frac{\tilde{\sigma}_{k,t+1|t}^2}{\sigma_{k,t+1|t}^2} \quad (40)$$

$$= \frac{1}{2} \ln \left( 1 + \sum_{i=0}^1 \sum_{j=0}^1 \text{prob}(s_{t+1} = j, s_t = i | I_t) (y_{k,t+1|t}^{(i,j)} - \tilde{y}_{k,t+1|t})^2 \right) \quad (41)$$

where  $y_{k,t+1|t}^{(i,j)} = E(y_{k,t+1} | s_{t+1} = j, s_t = i, I_t)$ , is the conditional mean of  $y_{k,t+1}$  given the path  $(s_t = i, s_{t+1} = j)$

$$y_{k,t+1|t}^{(i,j)} = \lambda_k \mu_j - \psi_k \lambda_k (y_{k,t} - \mu_i)$$

---

<sup>5</sup>Notice that the conditional variance of the one step ahead predictions,  $\sigma_{k,t+1|t}^2$ , given the path  $(s_t = i, s_{t+1} = j)$ ,  $i, j = 0, 1$  is the same for the four possible paths.

and  $\tilde{y}_{k,t+1|t}$  and  $\tilde{\sigma}_{k,t+1|t}^2$  are the mean and variance of the mixture, respectively, given by

$$\tilde{y}_{k,t+1|t} = \sum_{i=0}^1 \sum_{j=0}^1 \text{prob}(s_{t+1} = j, s_t = i | I_t) y_{k,t+1|t}^{(i,j)}, \quad (42)$$

$$\tilde{\sigma}_{k,t+1|t}^2 = \sigma_{k,t+1|t}^2 + \sum_{i=0}^1 \sum_{j=0}^1 \text{prob}(s_{t+1} = j, s_t = i | I_t) (y_{k,t+1|t}^{(i,j)} - \tilde{y}_{k,t+1|t})^2. \quad (43)$$

### 2.3 Mixing frequencies

Recently, Aruoba and Diebold (2010) extracted an economic activity index from a linear dynamic factor model that uses four monthly indicators, employment, industrial production, real personal income less transfers and real manufacturing and trade sales, along with a quarterly indicator, GDP. Although they explicitly left for further research the assessment of the possible presence of regime-switching in the factors that are extracted from dynamic factor models, the treatment of mixing frequencies within MS-DFM is not straightforward. This section analyzes the peculiarities of its implementation in the MS-DFM setup.

Quarterly series that refer to stocks can easily be converted into monthly observations since they simply refer to quantities which are measured at a particular time and do not require any time restriction. Accordingly, these series can be treated as observed the month that they are issued and as unobserved otherwise. However, flow variables as GDP are measured during some time periods and must be temporally aggregated. Within linear frameworks, Mariano and Murasawa (2003) describe a time aggregation of flow variables which is based on the notion that quarterly time series can be viewed as sums of underlying monthly series in the corresponding quarter. Let us assume that arithmetic means can be approximated<sup>6</sup> by geometric means.<sup>7</sup> Hence, quarter-on-quarter growth rates ( $g_t$ ) of quarterly series are weighted averages of the monthly-on-monthly past growth rates ( $x_t$ )

---

<sup>6</sup>Other approaches in the literature which try to skip the approximation are not exempt of problems. The exact nonlinear filter of Proietti and Moauro (2006) involves approximations in its own and the exact linear filter of Aruoba, Diebold and Scotti (2009) assumes all indicators to be polynomial trends.

<sup>7</sup>The approximation is not very restrictive in practice. For example a constant growth of 1% each month in a particular quarter (annual growth of more than 12%), would imply a difference between arithmetic and geometric means of less than 0.4 percentage points.

of the (assumed to be known) monthly series in levels

$$g_t = \frac{1}{3}x_t + \frac{2}{3}x_{t-1} + x_{t-2} + \frac{2}{3}x_{t-3} + \frac{1}{3}x_{t-4}. \quad (44)$$

This expression accounts for the two characteristics of mixed processes. The first characteristic is time dependence since the processes are linear combinations of present and past monthly random variables  $x_t$  that depend on present and lagged hidden discrete state variables  $s_t, \dots, s_{t-k}$ . The second characteristic is that the monthly series of quarterly growth rates  $g_t$  is not observed since it is only available once every three monthly outcomes.

The first challenge of time dependence is to determine the lag order  $k$  in  $s_{t-k}$  that is required to correctly specify the probability distribution function of the given variable. To simplify the analysis, let us assume that there is only one quarterly indicator which is the first component of the vector of economic indicators. Let us also assume that the quarterly indicator has a loading factor equal to one and that its idiosyncratic component is a white noise with variance  $\sigma_1^2$ . In this case, the monthly growth rates are  $x_t = f_t + u_{1t}$ . Finally, we simplify the analysis by assuming that  $f_t = \mu_{s_t} + a_t$ , with  $a_t \sim N(0, \sigma_a^2)$ , and that  $\mu_{s_t} = \mu_i$  if  $s_t = i$ ,  $i = 0, 1$ . Then,  $x_t|s_t$  is distributed according to

$$f(x_t|s_t = i) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_t - \mu_i)^2}{2\sigma^2}\right), \quad (45)$$

with  $i = 0, 1$ , and  $\sigma^2 = \sigma_a^2 + \sigma_1^2$ .<sup>8</sup> Therefore,  $x_t$  is a mixture of two Gaussian densities

$$f(x_t) = \sum_{i=1}^2 \pi_i f(x_t|s_t = i), \quad (46)$$

where  $\pi_i$  is the unconditional probability of being in state  $i$ .

According to (44), the common factor affects the underlying monthly series of quarterly GDP growth rates contemporaneously and lagged. Since the monthly GDP growth  $x_t$  can be in either of the two states  $i = 0, 1$  (expansion or recession), and the monthly series of quarterly growth  $g_t$  is a linear combination of  $x_t$  and four lags, the process needs to account for up to  $2^5 = 32$  different business cycle paths. For this purpose, we follow

---

<sup>8</sup>Again, we are assuming regime-independent variances. It is easy to skip this assumption.

Hamilton (1989), to define  $s_t^*$  as the Markov process that accounts for these paths

$$\begin{aligned}
s_t^* &= 1 \text{ if } s_t = 0, s_{t-1} = 0, \dots, s_{t-4} = 0, \\
s_t^* &= 2 \text{ if } s_t = 1, s_{t-1} = 0, \dots, s_{t-4} = 0, \\
s_t^* &= 3 \text{ if } s_t = 0, s_{t-1} = 1, \dots, s_{t-4} = 0, \\
&\vdots \\
s_t^* &= 32 \text{ if } s_t = 1, s_{t-1} = 1, \dots, s_{t-4} = 1.
\end{aligned}$$

The Gaussian conditional probability density function of  $g_t$  given  $s_t^*$  is given by

$$f(g_t | s_t^* = j) = \frac{1}{\sqrt{2\pi\sigma_*^2}} \exp\left(-\frac{(g_t - \mu(s_t^* = j))^2}{\sigma_*^2}\right), \quad (47)$$

where  $j = 1, 2, \dots, 32$ ,

$$\begin{aligned}
\mu(s_t^* = 1) &= \frac{1}{3}\mu_0 + \frac{2}{3}\mu_0 + \mu_0 + \frac{2}{3}\mu_0 + \frac{1}{3}\mu_0 = 3\mu_0, \\
\mu(s_t^* = 2) &= \frac{1}{3}\mu_1 + \frac{2}{3}\mu_0 + \mu_0 + \frac{2}{3}\mu_0 + \frac{1}{3}\mu_0 = \frac{1}{3}\mu_1 + \frac{8}{3}\mu_0, \\
&\vdots \\
\mu(s_t^* = 32) &= \frac{1}{3}\mu_1 + \frac{2}{3}\mu_1 + \mu_1 + \frac{2}{3}\mu_1 + \frac{1}{3}\mu_1 = 3\mu_1,
\end{aligned} \quad (48)$$

and

$$\sigma_*^2 = \frac{19}{9}\sigma^2. \quad (49)$$

Hence, the density of  $g_t$  is

$$f(g_t) = \sum_{j=1}^{32} \pi_j^* f(g_t | s_t^* = j), \quad (50)$$

where  $\pi_j^*$  is the unconditional probability of being in the  $j$ -th state out of the 32 different states.

The empirical applications of this model to infer real-time business cycle probabilities will suffer from the curse of dimensionality problem.<sup>9</sup> And the curse of dimensionality

---

<sup>9</sup>The curse of dimensionality problem is magnified in practice when not all the dynamics of the time series  $x_t$  rely on a changing mean  $\mu_{s_t}$ . In these cases, the common factor includes up to  $p$  lags,  $x_t$  depends on  $s_t, s_{t-1}, \dots, s_{t-p}$ , and  $g_t$  depends on  $s_t, \dots, s_{t-4}, \dots, s_{t-p-4}$ . This results in more complex algorithms that should account for up to  $2^{p+5}$  different business cycle paths.

problem is magnified in practice when not all the dynamics of the time series  $x_t$  rely on a changing mean  $\mu_{s_t}$ . In these cases, the common factor includes up to  $p$  lags,  $x_t$  depends on  $s_t, s_{t-1}, \dots, s_{t-p}$ , and  $g_t$  depends on  $s_t, \dots, s_{t-4}, \dots, s_{t-p-4}$ . This results in more complex algorithms that should account for up to  $2^{p+5}$  different business cycle paths.

One simplification strategy, which has been used by Chauvet (1998) in the context of MS-DFM that use only monthly indicators, consists on approximating the switching mean model by a switching intercept model. This strategy would simplify the algorithm considerably since it would need to account only for 32 different states regarding the lag length of the autoregressive process. In this paper, we go even further. In Section 3 we present the results of a simulation study where we assess the performance of approximating the model that uses 32 different states by a simplifying version that uses only 2 different states. i.e., we evaluate the performance of assuming the density of the quarterly indicator as if it only relied on the two initial states

$$f(g_t) = \sum_{i=1}^2 \pi_i f(g_t | s_t = i). \quad (51)$$

We conclude that this approximation is quite accurate in many empirical situations.

The second characteristic of mixed processes is that quarterly series exhibit two missing observations within each quarter. Since these missing data and the absence of the latest releases due to data publication delays can both be treated in a similar manner in the context of MS-DFM, the next subsection describes a method to fill in the gaps and to produce inferences of unobserved series by using a nonlinear Kalman filter.

## 2.4 Dealing with missing data

The presence of both ragged ends and mixing frequencies generate missing data in dynamic factor models. We extend the procedure described by Mariano and Murasawa (2003), which deals with missing observations in linear DFM, to the Markov-switching setup. Following these authors, our proposal is based on filling in the missing observations with random numbers that are extracted from a random variable whose distribution is independent of the model parameters and rewriting the measurement equation appropriately to get that the nonlinear Kalman filter skips the random numbers. We show that

the parameters that maximize the likelihood and the inferences about the business cycle states are achieved as if all the variables were observed.

Without loss of generality, we focus on dealing with missing data in the  $t$ -th observation of a quarterly indicator which is the first variable of a set of monthly indicators.<sup>10</sup> Let  $\alpha$  be the vector that includes all the unknown model parameters, and let the quarterly indicator  $g_t$  be the first component of the vector of time series  $\mathbf{y}_t$ , i.e.,  $y_{1,t} = g_t$ . Let us define the variable

$$y_{1,t}^+ := \left\{ \begin{array}{ll} y_{1,t} & \text{if } y_{1,t} \text{ is observable} \\ z_t & \text{otherwise} \end{array} \right\}, \quad (52)$$

where  $z_t$  is a random variable whose distribution is independent of  $\alpha$ , for instance,  $z_t \sim N(0, \sigma_z^2)$ . Let  $\beta$  be the vector of parameters associated to the density function of  $z_t$  and  $f(z_t; \beta)$  the density function of  $z_t$ . To skip the row corresponding to nonobserved variables from the Kalman recursion, the measurement equation simply discards the row corresponding to the first variable which is set to the random draw whenever it is not observed. Accordingly, no modifications of the nonlinear algorithm used to estimate MS-DFM are needed apart from considering the time varying Kalman filter to zero out the missing observations.

To check that the method works in the Markov-switching setup, let us show that the estimates of  $\alpha$  that maximize the likelihood function when missing data are replaced by  $z_t$  are the estimates that also maximize the likelihood function of the observed data. For this purpose, let  $\mathbf{y}_t^+ = (y_{1,t}^+, y_{2,t}, \dots, y_{N,t})'$  and consider the joint density function of  $(\mathbf{y}_1^+, \dots, \mathbf{y}_T^+)$ . Since  $f(z_t; \beta)$  does not depend on  $\alpha$ , the likelihood can be written as

$$f(\mathbf{y}_1^+, \dots, \mathbf{y}_T^+; \alpha, \beta) = f(\mathbf{y}_1, \dots, \mathbf{y}_T; \alpha) \prod_{t \in A} f(z_t; \beta), \quad (53)$$

where  $A \subset \{1, \dots, T\}$  is the subset of missing observations for  $y_{1,t}$ . Taking logs, it is very easy to see that regardless the values of  $z_t$ , the argument that maximizes the log of  $f(\mathbf{y}_1^+, \dots, \mathbf{y}_T^+; \alpha)$  must also maximize the log of  $f(\mathbf{y}_1, \dots, \mathbf{y}_T; \alpha)$ .

In addition, let us show that the filtered probabilities that are obtained from observed and missing data  $\{\mathbf{y}_t^+\}_{t=1}^T$ ,  $prob(s_t = i | I_t^+)$ , do not depend on  $z_t$  either. For this purpose,

<sup>10</sup>The procedure described in this section can be easily extended to other situations. Examples are several missing data at time  $t$  and missing data on the monthly indicators.

let us consider the simplifying restrictions that  $x_t = \lambda_1 f_t + u_{1t}$ , where  $f_t = \mu_{s_t} + a_t$ , and  $u_{1t}$  is white noise. In this case, the probability distribution can be approximated by a mixture of 32 Gaussian densities.<sup>11</sup> Let  $I_t^+$  be the information set generated by  $(\mathbf{y}_1^+, \dots, \mathbf{y}_t^+)$ . Let

$$prob(s_{t-1} = l, s_{t-2} = m, s_{t-3} = n, s_{t-4} = o | I_{t-1}^+) \quad (54)$$

be the set of state probabilities which are computed at  $t - 1$ .<sup>12</sup> If we denote the states of  $s_t^*$  that contain  $s_t = i$  as  $C = \{j \in \{1, \dots, 32\} | s_t = i\}$ , it is easy to obtain the filtered probability of a given state  $i$  by adding these probabilities

$$prob(s_t = i | I_t^+) = \sum_{j \in C} prob(s_t^* = j | I_t^+). \quad (55)$$

Each of these 16 terms is given by

$$prob(s_t^* = j | I_t^+) = \frac{f(\mathbf{y}_t^+, s_t^* = j | I_{t-1}^+)}{f(\mathbf{y}_t^+ | I_{t-1}^+)}. \quad (56)$$

Let us assume that  $y_{1,t}$  is observed. Although we will show that this is actually the case, let us also assume that  $prob(s_t^* = j | I_{t-1}^+)$ ,  $j = 1, \dots, 32$ , does not depend on  $\{z_1, \dots, z_{t-1}\}$ . In this case,  $\mathbf{y}_t^+ = \mathbf{y}_t$  and the posterior probabilities can be computed as

$$prob(s_t^* = j | I_t^+) = \frac{f(\mathbf{y}_t | s_t^* = j, I_{t-1}^+) prob(s_t^* = j | I_{t-1}^+)}{\sum_{k=1}^{32} f(\mathbf{y}_t | s_t^* = k, I_{t-1}^+) prob(s_t^* = k | I_{t-1}^+)}, \quad (57)$$

with

$$f(\mathbf{y}_t | s_t^* = j, I_{t-1}^+) = \frac{1}{\sqrt{(2\pi)^N |\Sigma_{t|t-1}^j|}} \exp\left\{-\frac{1}{2}(\mathbf{y}_t - \mathbf{y}_{t|t-1}^j)' \left(\Sigma_{t|t-1}^j\right)^{-1} (\mathbf{y}_t - \mathbf{y}_{t|t-1}^j)\right\}, \quad (58)$$

where  $\mathbf{y}_{t|t-1}^j$  is the one step ahead forecast of  $\mathbf{y}_t$  given information up to time  $t - 1$  if  $s_t^* = j$ , and  $\Sigma_{t|t-1}^j$  is its variance-covariance matrix. In addition, for each state  $s_t^* = j$

$$prob(s_t^* = j | I_{t-1}^+) = p_{lj} prob(s_{t-1} = l, s_{t-2} = m, s_{t-3} = n, s_{t-4} = o | I_{t-1}^+), \quad (59)$$

where  $j, l, m, n, o \in \{0, 1\}$ . Hence, plugging (58) and (59) in (57), the desired filtered probabilities of a particular regime are obtained by using (55). Finally, one can also

<sup>11</sup>The generalization to processes of larger orders is straightforward.

<sup>12</sup>The initial probabilities can be set to either the ergodic probabilities or to 1/16.

compute the next input needed by the algorithm as

$$\begin{aligned} \text{prob}(s_t = i, s_{t-1} = l, s_{t-2} = m, s_{t-3} = n | I_t^+) &= \\ &= \sum_{o=1}^2 p(s_t = i, s_{t-1} = l, s_{t-2} = m, s_{t-3} = n, s_{t-4} = o | I_t^+). \end{aligned} \quad (60)$$

for given  $i, l, m, n$ . Notice that  $\text{prob}(s_t^* = j | I_t^+)$  does not depend on  $z_t$ .

Assume now that  $y_{1,t}$  is not observed. In this case,  $\mathbf{y}_t^+ = (z_t, y_{2,t}, \dots, y_{N,t})' = (z_t, \mathbf{y}_t^{-'})'$ , where  $\mathbf{y}_t^-$  are the observed variables at  $t$ . Then, since  $z_t$  does not depend on the remaining variables in the model, the posterior probabilities can be computed as

$$\text{prob}(s_t^* = j | I_t^+) = \frac{f(\mathbf{y}_t^-, s_t^* = j | I_{t-1}^+) f(z_t)}{f(\mathbf{y}_t^- | I_{t-1}^+)} \quad (61)$$

$$= \frac{f(\mathbf{y}_t^- | s_t^* = j, I_{t-1}^+) \text{prob}(s_t^* = j | I_{t-1}^+) f(z_t)}{\sum_{k=1}^{32} f(\mathbf{y}_t^- | s_t^* = k, I_{t-1}^+) \text{prob}(s_t^* = k | I_{t-1}^+) f(z_t)} \quad (62)$$

$$= \frac{f(\mathbf{y}_t^- | s_t^* = j, I_{t-1}^+) \text{prob}(s_t^* = j | I_{t-1}^+)}{\sum_{k=1}^{32} f(\mathbf{y}_t^- | s_t^* = k, I_{t-1}^+) \text{prob}(s_t^* = k | I_{t-1}^+)}, \quad (63)$$

which does not depend on  $z_t$ .

### 3 Monte Carlo simulations

In this section, we set up several Monte Carlo experiments to study how the real time features of incoming data might affect the relative empirical performance of traditional MS-DFM, which are restricted to work with balanced panels of data, with respect to our extension of MS-DFM, which are able to deal with ragged ends and mixing frequencies.

For this purpose, we generate a total of  $M = 1000$  sets of  $N$  idiosyncratic components  $\mathbf{u}_t^m$  of length  $T' = T + J$ , where  $T = 200$  refers to the in-sample data, and  $J = 10$  refers to the forecasting period. In the simulations, the time series are generated with equal variances  $\sigma_i^2 = \sigma^2$ . The dynamics of these idiosyncratic time series are assumed to follow autoregressive processes of order one with autoregressive parameters equal to 0.3.

In addition, we generate  $M = 1000$  dummy variables  $b_t^m$  of zeroes and ones of length  $T'$  which are used to simulate different sequences of expansions ( $b_t^m = 0$ ) and recessions ( $b_t^m =$

1). To ensure that the dummies share the US business cycle properties, we assume that  $b_t^m$  follows Markov chains with  $p_{00} = 0.9$  and  $p_{11} = 0.7$ . According to the NBER Business Cycle Dating Committee, these transition probabilities coincide with the percentage of quarters classified as expansions that are followed by expansions and the percentage of quarters classified as recessions that are followed by recessions in the period 1959.3-2010.3. Then, we generate  $M = 1000$  common factors,  $f_t^m$ , that follow Markov-switching processes by using the business cycle sequences  $b_t^m$  to classify the business cycle states, by assuming within state means of  $\mu_0 = 1$  and  $\mu_1 = -1$ , and by setting  $\sigma_a^2 = 1$ . Finally, using loading factors equal to one for all the series, we add the idiosyncratic components to the switching mean factors to generate  $M = 1000$  sets of  $N$  time series  $\{\mathbf{y}_t^m\}_{t=1}^{T'}$ .

To examine the effects of dealing with ragged ends in computing the real-time business cycle inferences, let us assume that an analyst faces the inference problem by assuming that there is one publication lag in four out of the set of  $N$  indicators used in the analysis. To complete the analysis, the simulations are also computed when these four indicators exhibit two publication lags. To examine the effect of the quality of the indicators in the forecasting accuracy, the series are generated with the same but increasing idiosyncratic variance  $\sigma^2$  of 0.5, 1.5, and 4.5. To analyze the role of  $N$ , the total number of indicators used in the analysis is 5 and 7.

Let us assume that the analyst wants to infer the probability of recession at  $T + j$ , with  $j = 0, 1, \dots, J$ , under two different scenarios from the set of  $N$  indicators. The first scenario consists on using traditional MS-DFM to infer recession probabilities at  $T + j$  with the (as large as possible) amount of information disposable at  $T + j$ . In this case, the analyst faces two alternatives. The first alternative consists on forecasting from the latest available balanced panel of  $N$  indicators. Hence, she has to compute one-step ahead forecasts to obtain  $prob(s_{T+j} = 1 | I_{T+j-1})$  and two-step ahead forecasts to obtain  $prob(s_{T+j} = 1 | I_{T+j-2})$  from the set of  $N$  indicators when there are one and two periods of publication lags, respectively.

The second scenario consists on using our extension of MS-DFM that is able to deal with ragged ends. In this case, the inferences can be computed from the  $N$  indicators even when four of them are not available at  $T + j$ , i.e., the analyst can compute  $prob(s_{T+j} =$

$1|I_{T+j}^+$ ), where  $I_{T+j}^+$  refers to information provided by the set of  $N - 4$  promptly published indicators up to  $T + j$  and the 4 delayed indicators up to  $T + j - h$ , with  $h = 1, 2$ . In this case, the variance of the  $N - 4$  indicators that are published timely is 1.5, and the variance of the 4 delayed indicators is allowed to change from 0.5 to 1.5, and 4.5. The role of the number of indicators is examined by allowing  $N$  to change from 5 to 7.

For each  $m$ -th replica, we quantify the ability of these estimation procedures to detect the actual state of the business by computing the Forecasting Quadratic Probability Score (*FQPS*):

$$FQPS_i = \frac{1}{M} \sum_{m=1}^M \frac{1}{J} \sum_{j=1}^{T+J} (p_{T+j,i}^m - b_{t+j}^m)^2. \quad (64)$$

In this expression,  $i = I$  in the case of traditional MS-DFM that forecast from the latest available balanced panel and  $i = II$  in the case of our extension of MS-DFM that is able to deal with ragged ends. Hence, the measure is the average over the  $M$  replications of the mean squared deviation of the different types of inferences from the generated business cycles.

Table 1 displays the *FQPS* statistics when four indicators exhibit one and two (in parentheses) publication lags. According to the results outlined in Section 2, the table shows that using the incoming information as it is early available helps to increase the accuracy of the models. For example, let us focus on the case of computing inferences from  $N = 5$  indicators when four of them exhibit a one-period publication delay in the case  $\sigma^2 = 1.5$ . When one-step ahead probability forecasts are computed from the balanced set of indicators, the inferences exhibit *FQPS* of 0.167. However, the table shows that the most accurate business cycle inferences are obtained when the MS-DFM are allowed to deal with ragged ends since the *FQPS* falls to 0.133. In addition, the accuracy gains of accounting for ragged ends increase when the publication delay is two months (*FQPS* of 0.190 vs 0.133).

Notably, the sharp increases in the forecasting accuracy detected below are achieved by using only one timely published indicator. When the number of indicators increases, the inferences computed from the model that accounts for ragged ends also outperform the inferences computed from the model that computes probability forecasts from the complete

set of indicators ( $FQPS$  of 0.107 vs 0.163), especially when the indicators exhibit larger publication delays ( $FQPS$  of 0.107 and 0.187, respectively).

The entries displayed in Table 1 show that the ability to compute business cycle inferences from larger models crucially depends on the signal to noise ratio of the early information. The last row of the table shows that  $FQPS$  rises from 0.103 to 0.107 and 0.110 when the idiosyncratic variance increases from 0.5 to 1.5 and 4.5. In spite of this comment, it is worth pointing out that the relatively better performance of the MS-DFM that deals with ragged ends is maintained when the indicators exhibit larger signal-to-noise ratios. As the idiosyncratic variance increases, the corresponding  $FQPS$  of the models that cannot deal with ragged ends rises from 0.158 to 0.163 and 0.175.

Finally, to examine the role of using indicators with different frequencies to compute business cycle inferences, we develop the following Monte Carlo experiment. We generate  $M = 1000$  sets of  $N = 4$  “monthly” indicators of length  $T = 200$  and  $\sigma^2 = 4.5$ , and  $M = 1000$  business cycle dummies as described above. Besides, for each simulation we generate 1 “quarterly” indicator which is provided with the publication characteristics of quarterly data: the indicator is observed only once each three observations (each quarter only the last observation is available and the first two observations are missing data), and its quarterly growth rates are weighted averages of the monthly past growth rates as described in (44). The when the idiosyncratic variance of the quarterly indicator is allowed to change from 0.5 to 4.5.

To isolate the effect of mixing frequencies from the ragged ends effect, the time series are generated in this case under the assumption that there is no publication delay. Accordingly, the analysis is restricted to examine the in-sample accuracy of a MS-DFM that uses only the balanced panel of monthly indicators compared with a MS-DFM that uses both monthly and quarterly indicators. For this purpose, we compute the Quadratic Probability Score ( $QPS$ )

$$QPS = \frac{1}{M} \sum_{m=1}^M \frac{1}{T} \sum_{t=1}^T (p_t^m - b_t^m)^2, \quad (65)$$

which can be interpreted as an in-sample  $FQPS$ .

Table 2 suggests that two features of our analysis of mixing frequencies in MS-DFM

are noteworthy. First, the enlarged MS-DFM that uses four monthly indicators and one quarterly indicator uniformly outperforms the MS-DFM that is restricted to use the four monthly indicators only. The *QPS* falls from 0.157 to 0.143 when the quarterly indicator is as noisy as the four monthly indicators ( $\sigma^2 = 4.5$ ), and to 0.143 when the quarterly indicator is less noisy ( $\sigma^2 = 0.5$ ) than the monthly indicators. Second, approximating the time-consuming model that uses the actually required 32 states by a fast simplifying version that uses only the 2 initial states does not incur in large performance declines, especially when the quarterly indicator is as noisy as the four monthly indicators. In particular, although the 32-state model reduces the *QPS* by about 21% with respect to the 2-state approximation (from 0.153 to 0.121) when  $\sigma^2 = 0.5$ , the magnitude of the reduction falls to only about 5% when  $\sigma^2 = 4.5$ . Noticeably, the fast approximation provided by the 2-state model still perform better than the model that uses the monthly indicators only. In that sense, we can conclude that if monthly and quarterly indicators exhibit similar idiosyncratic noises then the fast approximation obtained from the 2-state model is able to capture the business cycle inferences with a similar accuracy than the 32-state model.

## 4 Empirical application

The purpose of this section is to examine the relative empirical performance of our modified MS-DFM which is able to deal with ragged ends and mixed frequencies with respect to traditional MS-DFM which are restricted to use balanced panels of data. For this purpose, we use an updated version of the data set early used in a linear context by Aruoba and Diebold (2010).

### 4.1 In-sample analysis

The four monthly indicators used in the empirical analysis are monthly industrial production index, nonfarm payroll employment, personal income less transfer payments and real manufacturing and trade sales. Although the latest available data set has been downloaded on January, 15th 2011, the balanced panel of four monthly indicators only includes

data from 1967.01 to 2010.11 since income is only available up to December 2010 and sales is only available up to November 2010.

Since the seminal proposal of Diebold and Rudebusch (1996), the behavior of these series is assumed to follow the comovements and asymmetries that were designated by Burns and Mitchell (1946) as the key business cycle features. Following their lines, we fit a MS-DFM to the balanced panel of one hundred times the change in the natural logarithm of these four macroeconomic variables.<sup>13</sup> The maximum likelihood estimates of this monthly model, which are displayed in the top panel of Table 3, show that the estimates of the signal-to-noise ratios agree with the magnitudes used in the simulation experiments. In particular, the top values of the simulated signal-to-noise ratios are achieved by industrial production, the medium values are achieved by employment, and the lowest values are achieved by sales and income. In addition, the estimates show that the factor loadings are positive and statistically significant. Hence, the indicators are positively correlated with the estimated common factor. In line with this statement, Figure 1 shows that the coincident index describes a behavior that closely agrees with the NBER-designated US business cycles.<sup>14</sup>

Notably, the maximum likelihood estimates reported in the top panel of Table 4 also show that the transition probabilities are very persistent ( $p_{00} = 0.98$ ,  $p_{11} = 0.85$ ) and that the within-state means are separate from each other ( $\mu_0 = 0.32$ ,  $\mu_1 = -2.00$ ). Figure 2, which plots the probabilities that the coincident indicator is in the negative growth rate based on currently available information along with shaded areas that represent periods dated as recessions by the NBER, shows that the filtered probabilities are in striking accord with the professional consensus as to the history of US business cycles. According to the theory and the Monte Carlo simulations, Table 4 shows that when the particular features of real-time inferences are omitted, the high quality of the indicators used in the model leads to a very good in-sample business cycle performance ( $QPS = 0.049$ ). According to the results suggested by Camacho et al. (2011), we do not expect to find

---

<sup>13</sup>According to Stock and Watson (1991), all the linear autoregressive processes are estimated with two lags. According to Camacho and Perez Quiros (2007), the nonlinear factor is estimated with no lags.

<sup>14</sup>In the empirical analysis, we take it as given that the NBER correctly identifies the dates of business cycle turning points.

large improvements in the historical (or in-sample) business cycle performance of enlarged models due to the high quality of the indicators already included in the model.

Recently, Aruoba and Diebold (2010) proposed a linear DFM to construct a real activity economic indicator which is based on these four monthly indicators along with quarterly real GDP. Our extension of MS-DFM allows us to obtain nonlinear estimates of this data set that contains business cycle indicators with monthly and quarterly frequencies. Interestingly, Table 3 shows that the maximum likelihood estimates that refer to the monthly series and the common factor are similar to the estimates obtained when GDP was excluded from the model.<sup>15</sup> The dynamics of the common factor, which is plotted in Figure 3, is also in close agreement with the dynamics of the estimated common factor obtained from the model that excluded GDP. In addition, the filtered probabilities that the coincident indicator is in the negative growth rate, which are plotted in Figure 4, also show a remarkably success in matching the NBER reference dates. The *QPS* of the model that uses quarterly GDP data is 0.048, which shows that the mixed frequency model exhibits slightly better overall performance than the model that uses monthly indicators only.

However, Table 4 also shows that it is remarkably to note that the reductions in *QPS* become of up to 5% (from 0.454 to 0.430) when the analysis is restricted to the first month after a phase shifts and the reductions reach to 20% (from 0.222 to 0.171) when the analysis refers to the first period of an expansion. Accordingly, we conclude that although the monthly indicators are quite informative about the historical US business cycle, the quarterly GDP issues are very helpful to improve upon the business cycle performance at business cycle phase shifts, especially to accurately detect the the exit of recessions.

## 4.2 Real-time analysis

The previous in-sample analysis has been conducted with data of the most recent vintage. However, the real-time data can be deceptively less helpful in monitoring the real activity than the in-sample evaluations developed in the previous section using finally revised datasets. On the one hand, it has been argued in the related literature (see, for example,

---

<sup>15</sup>According to the results of the Monte Carlo experiment, we include GDP in the approximated 2-state model.

Diebold and Rudebusch, 1991) that the good performance of the end-of-sample vintages in examining the empirical performance of econometric models may be spurious, in the sense that the data actually available in real time include economic time series that are subject to revision and that the economic relationships may change over time. In our case, the measures of production, employment and sales are typically subject to substantial revisions that sometimes occur years after the official figures are firstly released. On the other hand, the in-sample analysis does not allow the researchers to evaluate the effects of managing the lack of synchronicity that characterizes the daily flow of macroeconomic information in the early assessments of the economic developments.

To perform a more realistic assessment of the actual empirical reliability of the MS-DFM, we evaluate their real-time performance at tracking the US business cycles in real time through a data set that consists on vintages obtained from January 15, 1976 to February 15, 2011. That is, the inferences are computed at each month  $t$  over the past 35 years that covers the period December, 1976 to January, 2011 by using only the data that would have been available at the middle of the month that follows the particular month in which the inference is computed. This is accomplished by estimating the models on recursively increasing samples of data vintages and evaluating the evidence for a new turning point at the last month of each sample. Hence, the real-time analysis does not assume knowledge of data revisions that were not available at the time the model would have been used and has to manage with incomplete data sets at the time of each inference.

To facilitate understanding, let us recall the stylized publication calendar of the economic indicators used in the real-time analysis. At the end of month  $t$ , Industrial Production is published on the 15th of the month  $t + 1$ ; Non-farm employees is published on the 8th of the month  $t + 1$ , Real Personal income is published on the 27th of the month  $t + 1$ , Real Manufacturing and Trade Sales is published on the 27th of the month  $t + 2$ ,<sup>16</sup> and GDP, is published on the 15th of  $t + 2$ , whenever  $t$  is March, June, September or December. To simplify the real-time analysis, we consider that the real time inferences are computed on the 15th of each month, where employment and industrial production are available for the previous month. On this day, of month  $t + 1$ , we infer the probabilities

---

<sup>16</sup>The nominal indicator is published on the 14th of  $t + 2$ .

of being in a recession at  $t$ ,  $prob(s_t = 1|I_t)$ , with industrial production and employment up to  $t$ , personal income up to  $t - 1$ , and real sales up to  $t - 2$ .

According to our theoretical and Monte Carlo results, the business cycle probabilities are inferred from different alternative strategies. The first strategy, called strategy A, consists on computing inferences from traditional MS-DFM which can only account for balanced datasets. This implies that the model cannot use either quarterly series or the information provided by the early published indicators since the dataset must be restricted to finish at  $t - 2$ . Within this strategy, the inferences can be computed from one of the two following alternatives. In first alternative, called strategy A1, the inferences computed at  $t - 2$  are considered as the prevailing business cycle conditions for period  $t$ , i.e., the probabilities  $prob(s_t = 1|I_t)$  are approximated by  $prob(s_{t-2} = 1|I_{t-2})$ . In the second alternative, called strategy A2, the probabilities at  $t$  are computed by projecting the estimated probabilities for period  $t - 2$  to the current state by multiplying latest inferences by the transition matrix,  $prob(s_t = 1|I_{t-2})$ .

Strategies A1 and A2 clearly miss the extremely valuable information about the current business cycle that is provided by the early published indicators. In particular, these inferences miss the data of personal income at  $t - 1$ , and industrial production and employment at  $t - 1$  and  $t$ . To overcome this drawback, the business cycle inferences are computed in strategy B by using the extensions of MS-DFM proposed in this paper. Finally, the inferences are computed in strategy C by enlarging the model with GDP in the 2-state approximation of MS-DFM with mixed frequencies described in Section 2.

Figures 5, 6, 7, and 8, which plot the real-time filtered probabilities estimated from strategies A1, A2, B, and C, respectively, help us to assess the empirical performance of the different strategies in real time. According to expected, when the analysis is developed in real time the figures show a significant deterioration in the models performance with respect to the in-sample results. Although the in-sample filtered probabilities plotted in Figures 2 and 4, which are computed from finally revised data, provided unequivocal jumps in probabilities that marked the start and the end of the US business cycle phases, the real-time probabilities plotted in Figures 5 to 8 produce noisier and less accurate signals of the business cycle.

However, the figures also show that there is a significant improvement in business cycle forecasting accuracy when the MS-DFM is allowed to deal with ragged ends and to less extent to deal with mixing frequencies. To evaluate these forecasting improvements, Table 5 displays the results of the out-of-sample  $FQPS$  for the four different strategies. According to our theoretical and Monte Carlo results, strategies B and C provide much better forecasting accuracy than strategies A1 and A2, with a reduction in  $FQPS$  of more than 35% when it is computed for the entire sample used in the real-time analysis. The comparison of  $FQPS$  in the case of strategy B ( $FQPS = 0.066$ ) and strategy C ( $FQPS = 0.079$ ) reveals that the model that uses monthly indicators outperforms the model that is enlarged with GDP. However, as occurred in the in-sample analysis, the business cycle forecasting improvements of the model that mixes monthly and quarterly indicators come from its ability to identify the business cycle around turning points, especially when the analysis is concentrated on the beginning of expansions. In particular, using the quarterly GDP along with the fourth monthly indicators in the first month after the phase shifts to compute business cycle inferences reduces the  $FQPS$  from 0.684 to 0.612, and from 0.437 to 0.255 when the analysis is restricted to the first month after the troughs.

## 5 Concluding remarks

Real-time data usually display the feature of ragged ends, which means that end-of-sample observations of time series are missing and only released with a time-lag. The asynchronous publication releases limits the empirical benefits of Markov-switching dynamic factor models to monitor the day-to-day economic developments since models are restricted to deal with balanced data vintages and cannot manage all the relevant new releases as they arrive. In practice, the business cycle inferences computed from these models either are available only with a delay of several months, or they are computed as forecasts of past inferences.

From the point of view of monitoring the business cycle conditions, we show in the paper that there is no reason to be late or to throw away the relevant information provided by the latest figures of promptly issued indicators. We theoretically show that when

the economic indicators are carefully selected to have large signal-to-noise ratios in the Kalman filter used to compute business cycle inferences the increase in the business cycle identification accuracy becomes substantial.

The extension of dynamic factor models with regime switches proposed in this paper is the missing piece of this puzzle. Following the linear proposal by Mariano and Murasawa (2003), the method is based on a nonlinear Kalman filter to fill in the gaps of the non-synchronous flow of data releases in an efficient manner. By means of several Monte Carlo experiments, we quantify the magnitude of the accuracy improvements provided by our proposal over traditional methods. According to the theory, the improvements basically depend on the quality of the indicators used in the analysis.

In addition, traditional Markov-switching dynamic factor models cannot deal with business cycle indicators of different frequencies, typically monthly and quarterly. In this paper, we also show how to mix monthly and quarterly indicators to infer the business cycle phases. We show that quarterly data can in practice be treated as monthly data that exhibit missing monthly observations within each quarter. Accordingly, the nonlinear state-space framework proposed to deal with ragged ends can also be used to combine business cycle indicators of different frequencies. This allows us to consider our proposal as an extension of the linear method proposed by Aruoba and Diebold (2010).

In the empirical application considered in this paper we find that our theoretical findings are borne out. We use a real-time collection of data vintages which are updated monthly using only the information that would have been available at each month of the last 35 years. The vintages use the five key indicators that agree with the definition of a recession provided by the NBER Business Cycle Dating Committee, which defines a recession as “a significant decline economic activity spread across the economy, normally visible in real GDP, real income, employment, industrial production, and wholesale-retail sales”. We obtain substantial improvements of our extension of Markov-switching dynamic factor models which produce real-time business cycle probabilities that track the business cycle well, with pronounced drops corresponding to the NBER-designated recessions.

## 6 Appendix

### Proof of Proposition 1:

Recall that the conditional density of  $y_{k,t+1|t}$  is given by

$$f(y_{k,t+1|t}|s_{t+1} = i, I_t) = \frac{1}{\sqrt{2\pi\sigma_{t+1|t}^2}} \exp\left(-\frac{1}{2\sigma_{t+1|t}^2} \left(y_{k,t+1|t} - \lambda_k f_{t+1|t}^{(i)}\right)^2\right),$$

where  $i = 0, 1$ ,  $\sigma_{t+1|t}^2$  is the variance of the density function of  $y_{k,t+1|t}|s_{t+1} = i$ , which is the same in both states,  $I_t$  is the information set at  $t$  assumed state independent, and  $f_{t+1|t}^{(i)} = E(f_{t+1|t}|s_{t+1} = i, I_t)$ . Then

$$\begin{aligned} KL &= \int \ln \frac{f(y_{k,t}|s_{t+1} = 1, I_t)}{f(y_{k,t}|s_{t+1} = 0, I_t)} f(y_{k,t+1}|s_t = 1, I_t) dy_{k,t+1} \\ &= \frac{1}{2\sigma_{k,t+1|t}^2} \int \left( \left(y_{k,t+1|t} - \lambda_k f_{t+1|t}^{(0)}\right)^2 - \left(y_{k,t+1|t} - \lambda_k f_{t+1|t}^{(1)}\right)^2 \right) f(y_{k,t+1}|s_t = 1, I_t) dy_{k,t+1} \\ &= \frac{1}{2\sigma_{k,t+1|t}^2} \int \left[ 2y_{k,t+1|t} \lambda_k \left(f_{t+1|t}^{(1)} - f_{t+1|t}^{(0)}\right) + \lambda_k^2 \left( \left(f_{t+1|t}^{(0)}\right)^2 - \left(f_{t+1|t}^{(1)}\right)^2 \right) \right] f(y_{k,t+1}|s_t = 1, I_t) dy_{k,t+1} \\ &= \frac{1}{2\sigma_{k,t+1|t}^2} \left[ 2\lambda_k^2 f_{t+1|t}^{(1)} \left(f_{t+1|t}^{(1)} - f_{t+1|t}^{(0)}\right) + \lambda_k^2 \left( \left(f_{t+1|t}^{(0)}\right)^2 - \left(f_{t+1|t}^{(1)}\right)^2 \right) \right] \\ &= \frac{\lambda_k^2}{2\sigma_{k,t+1|t}^2} \left[ \left(f_{t+1|t}^{(1)}\right)^2 - 2f_{t+1|t}^{(1)} f_{t+1|t}^{(0)} + \left(f_{t+1|t}^{(0)}\right)^2 \right] \\ &= \frac{\lambda_k^2}{2\sigma_{k,t+1|t}^2} \left( f_{t+1|t}^{(1)} - f_{t+1|t}^{(0)} \right)^2. \end{aligned}$$

The desired expressions (19) to (21) are obtained noting that  $f_{t+1|t}^{(i)} = \mu_i$ , in cases (i) and (ii) but  $f_{t+1|t}^{(i)} = \alpha_i + \phi f_{t|t}$  in case (iii), and that  $\sigma_{k,t+1|t}^2 = \sigma_k^2$  in case (i),  $\sigma_{k,t+1|t}^2 = \sigma_k^2 + \sigma_a^2$  in case (ii), and  $\sigma_{k,t+1|t}^2 = \sigma_k^2 + \lambda_k^2(\phi^2 P_{t|t} + \sigma_a^2)$  in case (iii), where  $f_{t|t}$  and  $P_{t|t}$  are the mean and variance when estimating the common factor with information up to time  $t$ .

### Proof of Proposition 2:

To evaluate the information content of  $\mathbf{y}_{k,t+1|t}$  when it is a subvector  $m \times 1$  of observed variables notice that

$$\geq f(\mathbf{y}_{k,t+1|t}|s_{t+1} = i, I_t) = \frac{1}{(\sqrt{2\pi})^m |\boldsymbol{\Sigma}_{k,t+1|t}|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{y}_{k,t+1} - \boldsymbol{\Lambda}_k f_{t+1|t}^{(i)})' \boldsymbol{\Sigma}_{k,t+1|t}^{-1} (\mathbf{y}_{k,t+1} - \boldsymbol{\Lambda}_k f_{t+1|t}^{(i)})\right) \quad i =$$

where  $\Sigma_{k,t+1|t}$  is the variance of  $\mathbf{y}_{k,t}|s_{t+1} = i, I_t$  which for all factor model specifications we have assumed to be the same in both states  $i = 0, 1$  and  $f_{t+1|t}^{(i)} = E(f_{t+1}|s_{t+1} = i, I_t)$ . Using the notation in Cover and Thomas (2006) for the multiple integral, then

$$\begin{aligned}
KL &= \frac{1}{2} \int (\mathbf{y}_{k,t+1} - \mathbf{\Lambda}_k f_{t+1|t}^{(0)})' \Sigma_{k,t+1|t}^{-1} (\mathbf{y}_{k,t+1} - \mathbf{\Lambda}_k f_{t+1|t}^{(0)}) f(\mathbf{y}_{k,t+1}|s_t = 1, I_t) d\mathbf{y}_{k,t+1} \\
&- \frac{1}{2} \int (\mathbf{y}_{k,t+1} - \mathbf{\Lambda}_k f_{t+1|t}^{(1)})' \Sigma_{k,t+1|t}^{-1} (\mathbf{y}_{k,t+1} - \mathbf{\Lambda}_k f_{t+1|t}^{(1)}) f(\mathbf{y}_{k,t+1}|s_t = 1, I_t) d\mathbf{y}_{k,t+1} \\
&= \frac{1}{2} \int \left( 2\mathbf{y}'_{k,t+1} \Sigma_{k,t+1|t}^{-1} \mathbf{\Lambda}_k \left( f_{t+1|t}^{(1)} - f_{t+1|t}^{(0)} \right) \right) f(\mathbf{y}_{k,t+1}|s_t = 1, I_t) d\mathbf{y}_{k,t+1} \\
&+ \frac{1}{2} \int \left( \left( f_{t+1|t}^{(0)} \right)^2 - \left( f_{t+1|t}^{(1)} \right)^2 \right) \mathbf{\Lambda}'_k \Sigma_{k,t+1|t}^{-1} \mathbf{\Lambda}_k f(\mathbf{y}_{k,t+1}|s_t = 1, I_t) d\mathbf{y}_{k,t+1} \\
&= \frac{1}{2} \left( \mathbf{\Lambda}'_k \Sigma_{k,t+1|t}^{-1} \mathbf{\Lambda}_k 2f_{t+1|t}^{(1)} \left( f_{t+1|t}^{(1)} - f_{t+1|t}^{(0)} \right) \right) + \frac{1}{2} \left( \left( f_{t+1|t}^{(0)} \right)^2 - \left( f_{t+1|t}^{(1)} \right)^2 \right) \mathbf{\Lambda}'_k \Sigma_{k,t+1|t}^{-1} \mathbf{\Lambda}_k \\
&= \frac{1}{2} \left( f_{t+1|t}^{(0)} - f_{t+1|t}^{(1)} \right)^2 \mathbf{\Lambda}'_k \Sigma_{k,t+1|t}^{-1} \mathbf{\Lambda}_k \\
&= \frac{1}{2} (\alpha_0 - \alpha_1)^2 \mathbf{\Lambda}'_k \Sigma_{k,t+1|t}^{-1} \mathbf{\Lambda}_k
\end{aligned}$$

which is the desired expression.

### Proof of Proposition 3:

First, notice that for the models considered in case (iv)

$$f(y_{k,t+1}|s_{t+1} = j, s_t = i, I_t) = \frac{1}{\sqrt{2\pi\sigma_{k,t+1|t}^2}} \exp\left(-\frac{1}{2\sigma_{k,t+1|t}^2} \left(y_{k,t+1|t} - y_{k,t+1|t}^{(i,j)}\right)^2\right)$$

where  $y_{k,t+1|t}^{(i,j)} = E(y_{k,t+1}|s_{t+1} = j, s_t = i, I_t)$ , is given by

$$y_{k,t+1|t}^{(i,j)} = \lambda_k \mu_j - \phi \lambda_k (f_{t|t}^{(i)} - \mu_i)$$

and  $f_{t|t}^{(i)} = E(f_t|s_t = i, I_t)$ , is given by the "collapsing" method

$$f_{t|t}^{(i)} = \frac{\sum_{h=0}^1 \text{prob}(s_{t-1} = h, s_t = i|I_t) f_{t|t}^{(h,i)}}{\text{prob}(s_t = i|I_t)};$$

and  $\sigma_{k,t+1|t}^2$  is the variance of the conditional one ahead prediction errors of  $y_{k,t+1}$  given the path  $(s_t, s_{t+1}) = (i, j)$ . Notice that it is the same for all the conditional densities (2<sup>2</sup>). The Gaussian mixture can be approximated by a single Gaussian with mean and variance

given by  $\tilde{y}_{k,t+1|t}$  and  $\tilde{\sigma}_{k,t+1|t}^2$ , respectively

$$\begin{aligned}\tilde{y}_{k,t+1|t} &= \sum_{i=0}^1 \sum_{j=0}^1 \text{prob}(s_{t+1} = j, s_t = i | I_t) y_{k,t+1|t}^{(i,j)} \\ \tilde{\sigma}_{k,t+1|t}^2 &= \sigma_{k,t+1|t}^2 + \sum_{i=0}^1 \sum_{j=0}^1 \text{prob}(s_{t+1} = j, s_t = i | I_t) (y_{k,t+1|t}^{(i,j)} - \tilde{y}_{k,t+1|t})^2.\end{aligned}\quad (66)$$

Then

$$\begin{aligned}KL(iv) &= \int f(y_{k,t+1} | s_{t+1} = j, s_t = i, I_t) \ln \frac{f(y_{k,t+1} | s_{t+1} = j, s_t = i, I_t)}{f(y_{k,t+1} | I_t)} dy_{k,t+1} \\ &= \frac{1}{2} \int f(y_{k,t+1} | s_{t+1} = j, s_t = i, I_t) \ln \frac{\tilde{\sigma}_{k,t+1|t}^2}{\sigma_{k,t+1|t}^2} dy_{k,t+1} \\ &+ \frac{1}{2} \int f(y_{k,t+1} | s_{t+1} = j, s_t = i, I_t) \left( \frac{(y_{k,t+1|t} - \tilde{y}_{k,t+1|t})^2}{\tilde{\sigma}_{k,t+1|t}^2} - \frac{(y_{k,t+1|t} - y_{k,t+1|t}^{(i,j)})^2}{\sigma_{k,t+1|t}^2} \right) dy_{k,t+1} \\ &= \frac{1}{2} \ln \frac{\tilde{\sigma}_{k,t+1|t}^2}{\sigma_{k,t+1|t}^2} \\ &= \frac{1}{2} \ln \frac{\sigma_{k,t+1|t}^2 + \sum_{i=0}^1 \sum_{j=0}^1 \text{prob}(s_{t+1} = j, s_t = i | I_t) (y_{k,t+1|t}^{(i,j)} - \tilde{y}_{k,t+1|t})^2}{\sigma_{k,t+1|t}^2} \\ &= \frac{1}{2} \ln \left( 1 + \frac{\sum_{i=0}^1 \sum_{j=0}^1 \text{prob}(s_{t+1} = j, s_t = i | I_t) (y_{k,t+1|t}^{(i,j)} - \tilde{y}_{k,t+1|t})^2}{\sigma_{k,t+1|t}^2} \right).\end{aligned}\quad (67)$$

since the integral of the square differences to the mean of the process weighted by the density function is, by definition, the variance of the distribution, that divided by itself is equal to 1. Therefore the integral in (67) is just zero. The desired result is then obtained substituting the variance of the mixture  $\tilde{\sigma}_{k,t+1|t}^2$  by its expression given in (66).

**Proposition 4:**

Notice that in case (v) the  $k$  observed series is given by

$$y_{k,t+1} = \lambda_k f_{t+1} + u_{k,t+1}$$

where

$$\begin{aligned}f_{t+1} &= \mu_{s_{t+1}} + a_{t+1} \\ u_{k,t+1} &= \psi_k u_{k,t} + \epsilon_{k,t+1}.\end{aligned}$$

For stationary AR(1) processes, we can rewrite  $u_{k,t+1} = \frac{\epsilon_{k,t+1}}{1-\psi_k L}$  where  $L$  is the lag operator such that  $Lx_t = x_{t-1}$ . Therefore

$$\begin{aligned} y_{k,t+1} &= \lambda_k f_{t+1} + \frac{\epsilon_{k,t+1}}{1-\psi_k L} \\ &= \lambda_k f_{t+1} - \lambda_k \psi_k f_t + \psi_k y_{k,t} + \epsilon_{k,t+1} \end{aligned}$$

and  $y_{k,t+1|t}^{(i,j)} = E(y_{k,t+1}|s_{t+1} = j, s_t = i, I_t) = \lambda_k \mu_j - \lambda_k \psi_k \mu_i + \psi_k y_{k,t}$ . Proceeding as in proposition 3, the Gaussian mixture can be approximated by a single Gaussian with mean and variance given by  $\tilde{y}_{k,t+1|t}$  and  $\tilde{\sigma}_{k,t+1|t}^2$ , respectively in Proposition 4. Then

$$\begin{aligned} KL(v) &= \int f(y_{k,t+1}|s_{t+1} = j, s_t = i, I_t) \ln \frac{f(y_{k,t+1}|s_{t+1} = j, s_t = i, I_t)}{f(y_{k,t+1}|I_t)} dy_{k,t+1} \\ &= \frac{1}{2} \int f(y_{k,t+1}|s_{t+1} = j, s_t = i, I_t) \ln \frac{\tilde{\sigma}_{k,t+1|t}^2}{\sigma_{k,t+1|t}^2} dy_{k,t+1} \\ &+ \frac{1}{2} \int f(y_{k,t+1}|s_{t+1} = j, s_t = i, I_t) \left( \frac{(y_{k,t+1|t} - \tilde{y}_{k,t+1|t})^2}{\tilde{\sigma}_{k,t+1|t}^2} - \frac{(y_{k,t+1|t} - y_{k,t+1|t}^{(i,j)})^2}{\sigma_{k,t+1|t}^2} \right) dy_{k,t+1} \\ &= \frac{1}{2} \ln \frac{\tilde{\sigma}_{k,t+1|t}^2}{\sigma_{k,t+1|t}^2} \\ &= \frac{1}{2} \ln \frac{\sigma_{k,t+1|t}^2 + \sum_{i=0}^1 \sum_{j=0}^1 \text{prob}(s_{t+1} = j, s_t = i|I_t) (y_{k,t+1|t}^{(i,j)} - \tilde{y}_{k,t+1|t})^2}{\sigma_{k,t+1|t}^2} \\ &= \frac{1}{2} \ln \left( 1 + \frac{\sum_{i=0}^1 \sum_{j=0}^1 \text{prob}(s_{t+1} = j, s_t = i|I_t) (y_{k,t+1|t}^{(i,j)} - \tilde{y}_{k,t+1|t})^2}{\sigma_{k,t+1|t}^2} \right). \end{aligned}$$

## References

- [1] Aruoba, B., Diebold, F., and Scotti, C. 2009. Real-time measurement of business conditions. *Journal of Business and Economic Statistics* 27: 417-427.
- [2] Aruoba, B., and Diebold, F. 2010. Real-time macroeconomic monitoring: Real activity, inflation, and interactions. *American Economic Review: Papers and Proceedings* 100: 20-24.
- [3] Burns, A., and Mitchell, W. 1946. Measuring business cycles. National Bureau of Economic Research, New York.
- [4] Camacho, M., and Perez Quiros, G. and Poncela P. 2011. Extracting nonlinear signals from several economic indicators. Mimeo.
- [5] Camacho, M., and Perez Quiros, G. 2007. Jump-and-rest effect of U.S. business cycles. *Studies in Nonlinear Dynamics and Econometrics* 11(4): article 3.
- [6] Chauvet, M. 1998. An econometric characterization of business cycle dynamics with factor structure and regime switches. *International Economic Review* 39: 969-96.
- [7] Chauvet, M., and Hamilton, J. 2006. Dating business cycle turning points in real time. In *Nonlinear Time Series Analysis of Business Cycles*, eds. C. Milas, P. Rothman, and D. Van Dijk. Amsterdam: Elsevier Science, pp. 1-54.
- [8] Chauvet, M., and Piger, J. 2005. The real-time performance of business cycle dating methods. University of California working paper.
- [9] Chauvet, M., and Piger, J. 2008. A comparison of the real-time performance of business cycle dating methods. *Journal of Business and Economic Statistics* 26: 42-49.
- [10] Cover, T. and Thomas, J. 2006. *Elements of information theory*. Wiley, New Jersey.
- [11] Diebold, F., and Rudebusch, G. 1991. Forecasting output with the composite leading index: A real-time analysis. *Journal of the American Statistical Association* 86: 603-610.

- [12] Diebold, F., and Rudebusch, G. 1996. Measuring business cycles: A modern perspective. *Review of Economics and Statistics* 78: 67-77.
- [13] Hamilton, J. 1989. A new approach to the economic analysis of nonstationary time series and the business cycles. *Econometrica* 57: 357-384.
- [14] Hamilton, J. 2011. Calling recessions in real time. *International Journal of Forecasting* 27: 1006-1026.
- [15] Harrison, P. J. and Stevens, C.F. 1976. Bayesian forecasting. *Journal of the Royal Statistical Society, Ser. B* 38: 205-247.
- [16] Hershey, J.R. and Olsen, P. A. 2007. Approximating the Kullback-Leibler divergence between Gaussian mixture models, in: IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2007, vol.4, pp. IV-317–IV-320.
- [17] Kim, C. 1994. Dynamic linear models with Markov-switching. *Journal of Econometrics* 60: 1-22.
- [18] Kim, C., and Nelson, C. 1998. Business cycle turning points, a new coincident index, and tests of duration dependence based on a dynamic factor model with regime switching. *Review of Economics and Statistics* 80: 188-201.
- [19] Kim, C., and Yoo, J.S. 1995. New index of coincident indicators: A multivariate Markov switching factor model approach. *Journal of Monetary Economics* 36: 607-630.
- [20] Mariano, R., and Murasawa, Y. 2003. A new coincident index os business cycles based on monthly and quarterly series. *Journal of Applied Econometrics* 18: 427-443.
- [21] Peña, D., and Guttman, I. 1989. Optimal collapsing of mixture distributions in Robust Recursive Estimation. *Communications in Statistics, Theory and Methods* 18: 817-834.
- [22] Proietti, T. and Moauro, F. 2006. Dynamic factor analysis with non-linear temporal aggregation constraints. *Appl. Statist.* 55: 281-300.

- [23] Stock, J., and Watson, M. 1991. A probability model of the coincident economic indicators. In *Leading Economic Indicators: New Approaches and Forecasting Records*, edited by K. Lahiri and G. Moore, Cambridge University Press.

Table 1. Analysis of ragged ends in MS-DFM

$\sigma^2$	Balanced panels		Unbalanced panels	
	$N$			
	5	7	1+4	3+4
0.5	0.159 (0.186)	0.158 (0.184)	0.125 (0.126)	0.103 (0.104)
1.5	0.167 (0.190)	0.163 (0.187)	0.133 (0.133)	0.107 (0.107)
4.5	0.201 (0.202)	0.175 (0.193)	0.145 (0.145)	0.110 (0.110)

Notes.  $N$  is the number of indicators  $\sigma^2$  is the variance of their idiosyncratic components. In balanced MS-DFM, entries show the average over the replications of the averaged squared deviation of one- and two- (in parentheses) step-ahead filtered probabilities of low-mean state from the 1000 generated business cycle sequences. In MS-DFM with unbalanced panels, 1 and 3 variables with variance 1.5 are assumed to be timely available when the inference is computed and 4 indicators with variance 0.5, 1.5 and 4.5 are published with one- and two-month lags.

Table 2. Analysis of mixing frequencies

Number of states	4 monthly and 1 quarterly		4 monthly
	Variance of quarterly series		
	$\sigma^2=0.5$	$\sigma^2=4.5$	
2 states	0.153	0.154	0.157
32 states	0.121	0.143	No need

Notes. The variance of the monthly idiosyncratic components is equal to 4.5. The variance of the quarterly idiosyncratic variable is  $\sigma^2 = 0.5$  and 4.5. We also consider the possibility of not including the quarterly variable. The entries show the average over the replications of the averaged squared deviation of in-sample filtered probabilities of low-mean state from the 1000 generated business cycle sequences.

Table 3. Maximum likelihood estimates

Monthly						
		IP	Empl	Inc	Sales	GDP
Indicators	$\lambda_i$	0.69 (0.03)	0.42 (0.02)	0.28 (0.04)	0.46 (0.03)	
	$\phi_1$	-0.18 (0.08)	0.24 (0.03)	-0.20 (0.02)	-0.34 (0.04)	
	$\phi_2$	-0.16 (0.08)	0.54 (0.04)	-0.05 (0.04)	-0.15 (0.05)	
	$\sigma_i^2$	0.26 (0.04)	0.27 (0.02)	0.85 (0.03)	0.57 (0.03)	
Factor	$\mu_1$	$\mu_2$	$\sigma_{a^*}^2$	$p_{00}$	$p_{11}$	
	0.32 (0.07)	-2.00 (0.20)	1	0.98 (0.01)	0.85 (0.05)	
Monthly and quarterly						
		IP	Empl	Inc	Sales	GDP
Indicators	$\lambda_i$	0.67 (0.03)	0.42 (0.02)	0.29 (0.04)	0.48 (0.03)	0.30 (0.02)
	$\phi_1$	-0.07 (0.07)	0.24 (0.03)	-0.21 (0.02)	-0.37 (0.04)	0.024 (0.40)
	$\phi_2$	-0.07 (0.07)	0.54 (0.04)	-0.06 (0.04)	-0.17 (0.05)	-0.57 (0.22)
	$\sigma_i^2$	0.55 (0.04)	0.51 (0.02)	0.91 (0.03)	0.73 (0.03)	0.47 (0.15)
Factor	$\mu_1$	$\mu_2$	$\sigma_{a^*}^2$	$p_{00}$	$p_{11}$	
	0.29 (0.07)	-2.00 (0.21)	1	0.98 (0.01)	0.83 (0.06)	

Table 4. In-sample (1967.01-2010.11) empirical performance

	MS-DFM M	MS-DFM M+Q
Total	0.049	0.048
Turning points	0.454	0.430
Troughs	0.222	0.171

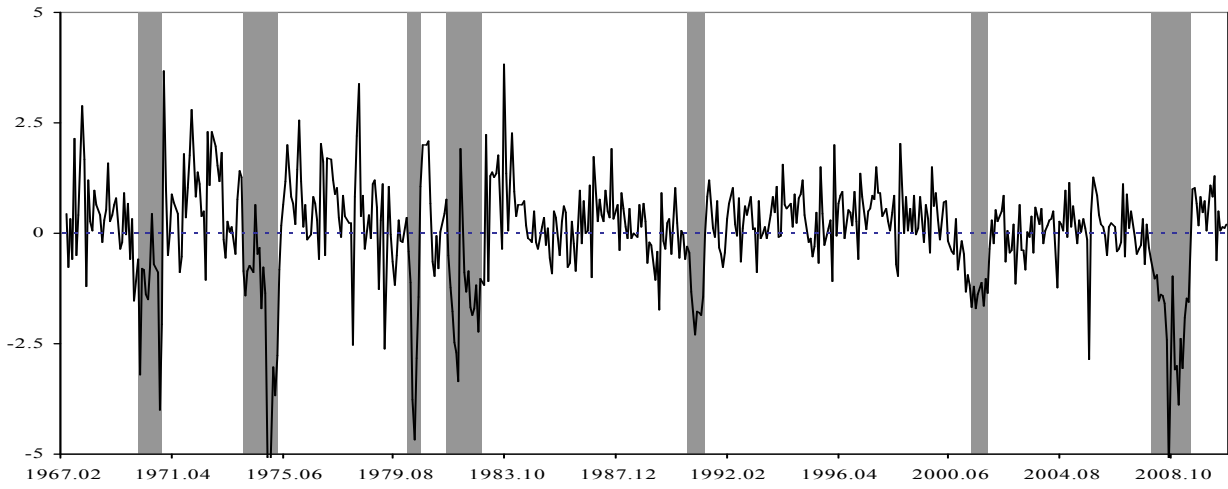
Note. Entries labelled as “total” refer to *QPS* statistics. In the case of entries labelled as “turning points”, the *QPS* is computed by using the first month after the phase shifts. In the case of entries labelled as “troughs”, the *QPS* is computed by using the first month of the expansions.

Table 5. Real-time (1976.10-2010.11) empirical performance

	MS-DFM Strategy A1	MS-DFM Strategy A2	MS-DFM Strategy B	MS-DFM Strategy C
Total	0.099	0.100	0.066	0.079
Turning points	0.783	0.675	0.684	0.612
Troughs	0.627	0.428	0.437	0.255

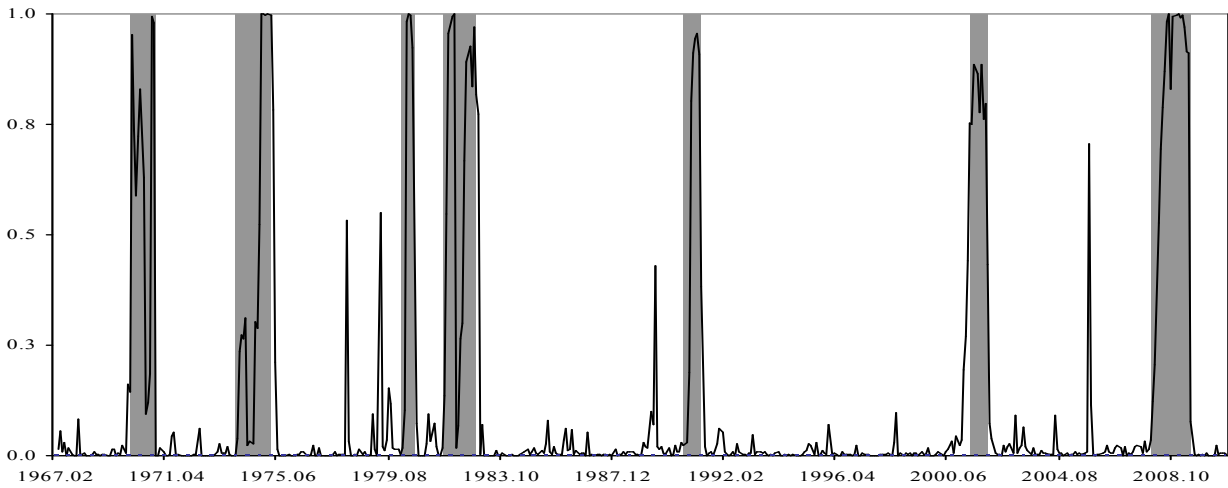
Note. Entries labelled as “total” refer to *FQPS* statistics. In the case of entries labelled as “turning points”, the *FQPS* is computed by using the first month after the phase shifts. In the case of entries labelled as “troughs”, the *FQPS* is computed by using the first month of the expansions. The forecasting strategies are defined in the text.

Figure 1. In-sample common factor from 4 monthly



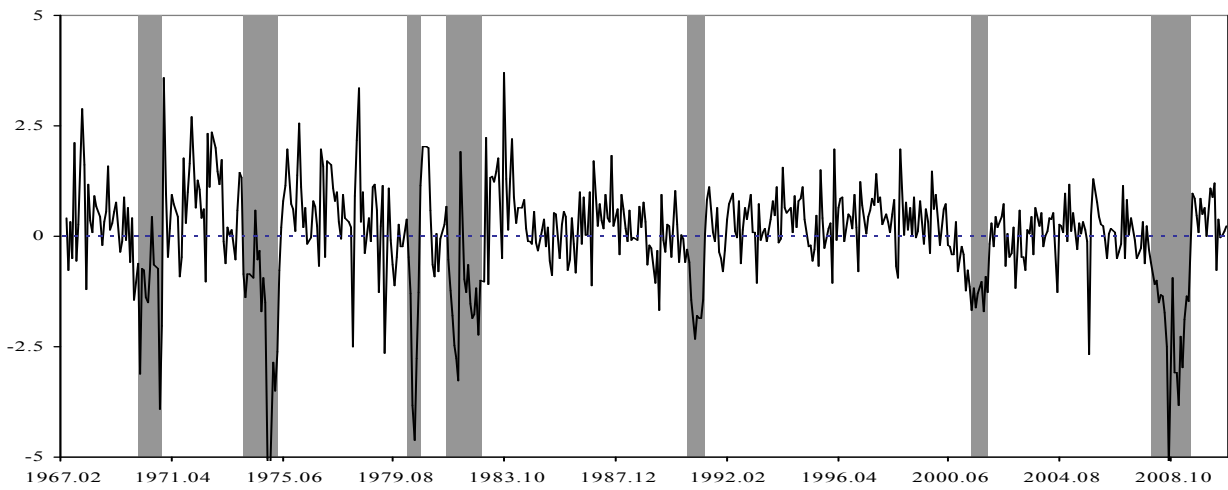
Notes. Shaded areas correspond to recessions as documented by the NBER.

Figure 2. In-sample filtered recession probabilities from 4 monthly



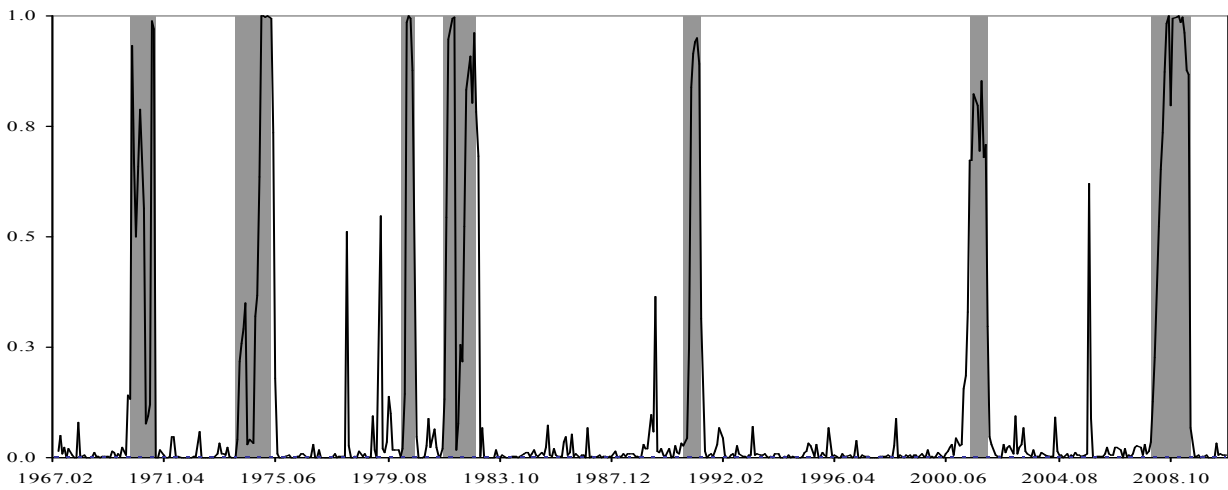
Notes. Shaded areas correspond to recessions as documented by the NBER.

Figure 3. In-sample common factor from 4 monthly and 1 quarterly



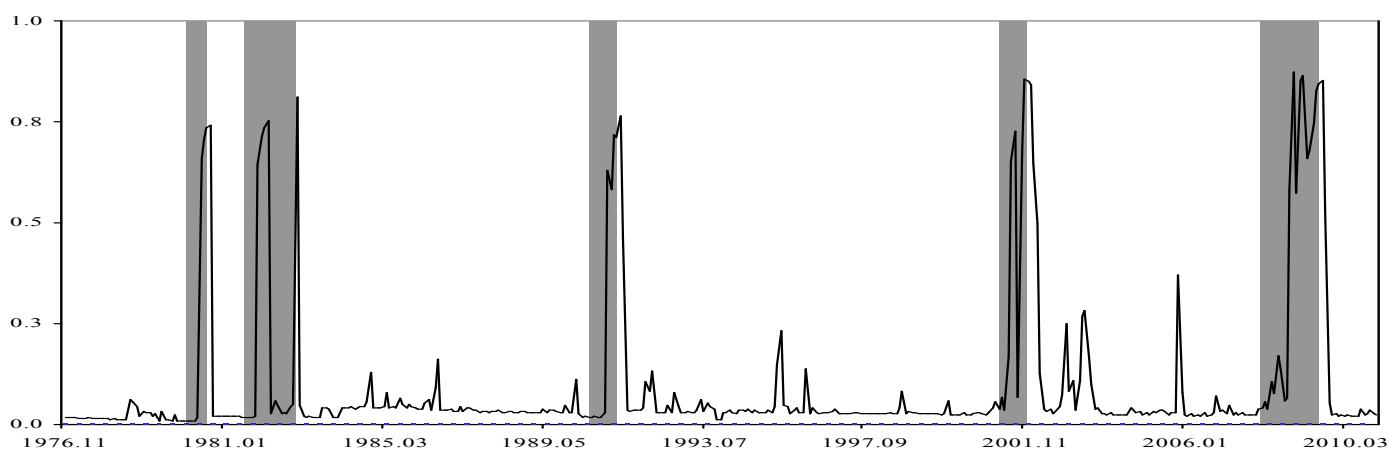
Notes. Shaded areas correspond to recessions as documented by the NBER.

Figure 4. In-sample filtered recession probabilities from 4 monthly and 1 quarterly



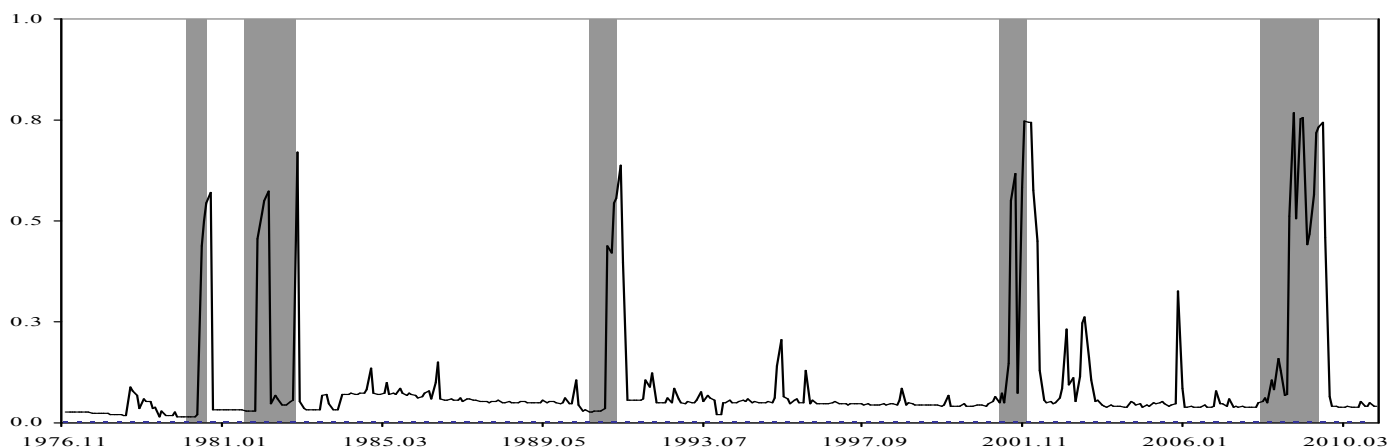
Notes. Shaded areas correspond to recessions as documented by the NBER.

Figure 5. Real time probabilities at  $t-2$  (plotted in  $t$ ) with information of 4 monthly up to  $t-2$



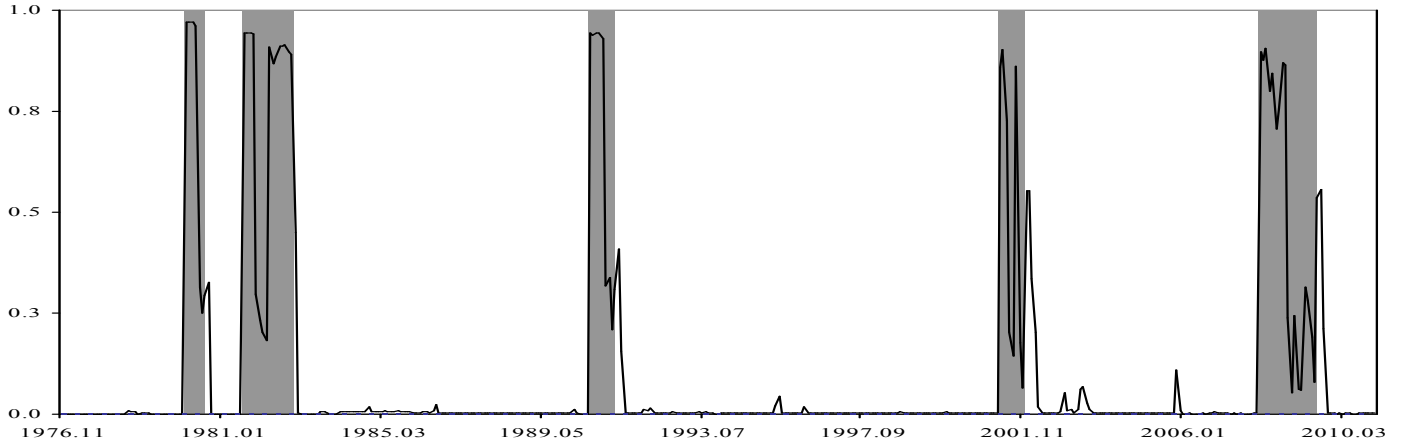
Notes. Shaded areas correspond to recessions as documented by the NBER. The figure plots the probabilities of recession in real time in period  $t$  using a balanced panel with published information in  $t-2$  and using those probabilities as if they were probabilities in period  $t$ . This is called in the text strategy A2.

Figure 6. Real time probabilities at  $t$  with information of 4 monthly up to  $t-2$



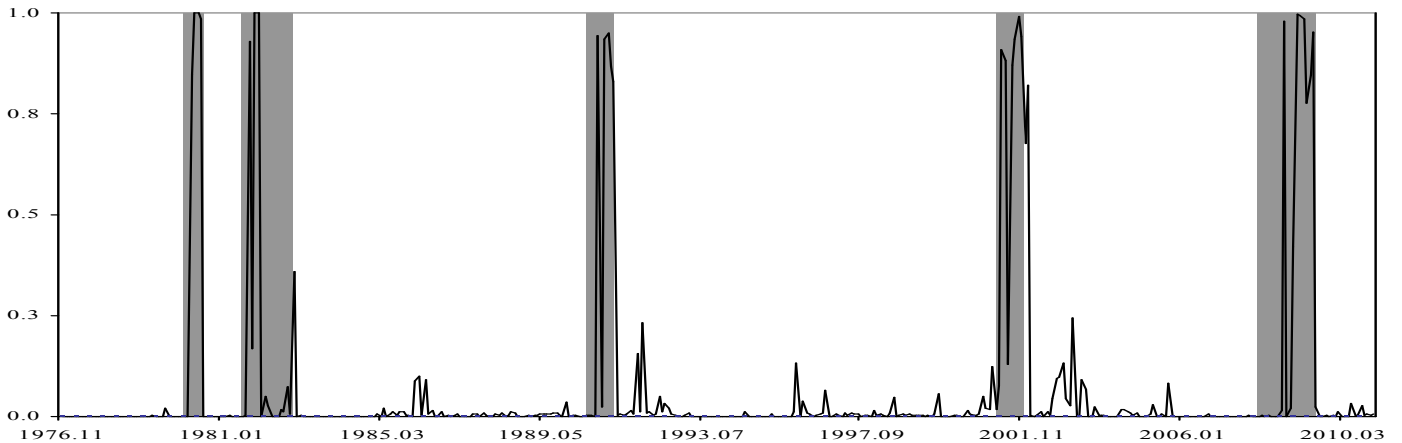
Notes. Shaded areas correspond to recessions as documented by the NBER. The figure plots the probabilities of recession in real time in period  $t$  using a balanced panel with published information in  $t-2$  and inferring, using the transition probability matrix, the probability of being in recession in  $t$ . This is called in the text strategy A1.

Figure 7. Real time probabilities at  $t$  with information of 4 monthly up to  $t$



Note: Shaded areas correspond to recessions as documented by the NBER. The figure plots the probabilities of recession in real time in period  $t$  using an unbalanced panel with published information in  $t$  and inferring, using the transition probability matrix updated with partial information,  $\pi_t$ , the probability of being in recession in  $t$ . This is called in the text strategy B.

Figure 8. Real time probabilities at  $t$  with information of 4 monthly up to  $t$  plus GDP series



Notes. Shaded areas correspond to recessions as documented by the NBER. The figure plots the probabilities of recession in real time in period  $t$  using an unbalanced panel with published information in  $t$ , including GDP series, and inferring, using the transition probability matrix updated with partial information,  $\pi_t$ , the probability of being in recession in  $t$ . This is called in the text strategy C.