DOCUMENT

DE TRAVAIL

N° 384

FORECASTING GDP OVER THE BUSINESS CYCLE IN A MULTI-FREQUENCY AND DATA-RICH ENVIRONMENT

Marie Bessec and Othman Bouabdallah

June 2012

BANQUE DE FRANCE
EUROSYSTÈME

DIRECTION GÉNÉRALE DES ÉTUDES ET DES RELATIONS INTERNATIONALES

DIRECTION GÉNÉRALE DES ÉTUDES ET DES RELATIONS INTERNATIONALES

FORECASTING GDP OVER THE BUSINESS CYCLE IN A MULTI-FREQUENCY AND DATA-RICH ENVIRONMENT

Marie Bessec and Othman Bouabdallah

June 2012

Forecasting GDP over the business cycle in a multi-frequency and data-rich environment

Marie Bessec[1]

Othman Bouabdallah[2]

[1]Banque de France, DGEI-DCPM-DIACONJ et LEDa-Université Paris Dauphine.

[2]Banque de France, DGEI-DCPM-FIPU, email: othman.bouabdallah@banque-france.fr

We thank Anindya Banerjee, Laurent Ferrara, Christian Francq, Ana Beatrix Galvão and Sheheryar Malik for their comments and suggestions, as well as Michel Juillard for his help with computational issues. We also thank Marie-Pierre Hourié-Felske for excellent research assistance. This paper reflects the opinions of the authors and does not necessarily express the views of the Banque de France.

Résumé : Dans cet article, nous introduisons un nouveau modèle de régression : le modèle MS-FaMIDAS. Cette spécification répond à plusieurs besoins du prévisionniste. Elle permet d'exploiter l'information dans une grande base de données de fréquence plus élevée que la variable à prévoir. Elle autorise par ailleurs des changements de régime dans la relation entre la variable endogène et ses prédicteurs. Des simulations de Monte Carlo montrent que cette spécification présente, en échantillon et hors échantillon, une qualité d'ajustement satisfaisante et prédit avec succès les changements de régime. Nous appliquons ensuite ce nouveau modèle au taux de croissance du PIB américain de 1959 à 2010 et détectons correctement les récessions américaines en exploitant le lien entre le taux de croissance du PIB et des variables financières mensuelles.

Mots-clés : changement de régime markovien, facteurs, MIDAS, prévision du PIB
Code JEL : C22, E32, E37.

Abstract: This paper merges two specifications developed recently in the forecasting literature: the MS-MIDAS model introduced by Guérin and Marcellino [2011] and the MIDAS-factor model considered in Marcellino and Schumacher [2010]. The MS-factor MIDAS model (MS-FaMIDAS) that we introduce incorporates the information provided by a large data-set, takes into account mixed frequency variables and captures regime-switching behaviors. Monte Carlo simulations show that this new specification tracks the dynamics of the process quite well and predicts the regime switches successfully, both in sample and out-of-sample. We apply this new model to US data from 1959 to 2010 and detect properly the US recessions by exploiting the link between GDP growth and higher frequency financial variables.

Keywords: Markov-Switching, factor models, mixed frequency data, GDP forecasting.
JEL classification: C22, E32, E37.

Introduction

The recent financial crisis has intensified among practitioners the interest in models differentiating GDP dynamics over the course of the business cycle, as firstly initiated by Hamilton [1989]. In order to forecast the GDP dynamics, macroeconomists can mobilize a very large set of indicators as Stock and Watson [2005] suggest. In this context, using common factors reflecting the comovements of these indicators is proved to be a convenient way to summarize this information. These indicators are very often available at higher frequencies than the targeted variable (GDP). This aggregation issue is quite successfully treated by MIxed DAta Sampling (MIDAS) models introduced by Ghysels, Santa-Clara and Valkanov [2004] and Ghysels, Sinko and Valkanov [2007]. This paper is at the crossroad of these three strands of the literature.

The MIDAS models are regressions involving variables sampled at different frequencies. In this framework, a low frequency variable can be explained by higher frequency indicators without any time aggregation procedure. A distributed lagged function can be used to get a parsimonious specification of the relationship between the dependent variable and the higher frequency variables. While MIDAS models have been first applied to financial data[3], they became a popular tool to forecast macroeconomic variables such as GDP growth as well. Forecasters use specifications relating the GDP variable to a handful of monthly leading indicators or rely on combinations of MIDAS models to deal with the potentially large number of indicators.[4] See Andreou, Ghysels and Kourtellos [2010] for a survey of this literature.

There are two recent extensions which are designed to forecast macroeconomic variables: MIDAS factor models by Marcellino and Schumacher [2010] and Markov-Switching MIDAS models by Guérin and Marcellino [2011]. In addition to involving mixed frequency data, the first class of models allows the use of information provided by a large dataset and can handle unbalanced samples that practitioners usually face due to different publication lags. The second class incorporates regime changes in the parameters of the relationship between the low and high frequency variables. Moreover, it gives qualitative information about the state of the economy. This provides a useful tool for the business cycle analysis.

[3]See Ghysels, Santa-Clara and Valkanov [2005], Ghysels, Santa-Clara and Valkanov [2006] and Ghysels et al. [2007] for applications to equity returns, Clements, Galvao and Kim [2008] to exchange rates.

[4]For instance, Clements and Galvao [2008], Clements and Galvao [2009], Bai, Ghysels and Wright [2009], Armesto, Hernández-Murillo, Owyang and Piger [2010b], Armesto, Engemann and Owyang [2010a] apply this model to forecast US GDP, Kuzin, Marcellino and Schumacher [2011] GDP in the euro area, Foroni, Marcellino and Schumacher [2011] euro area and US GDP.

In this paper, we introduce the MS Factor MIDAS model which captures both comovements and regime shifts in the dynamics of the variables and is implementable on mixed frequency data. We consider the dynamic factor model of Giannone, Reichlin and Small [2008] estimated with the 2-step method of Doz, Giannone and Reichlin [2011]. This approach can deal with the unbalanced data availability at the end of the sample due to uneven publication lags. Note that we allow a switch on the coefficients of the equation of the dependent variable (the coefficients of the GDP equation in our application) like Guérin and Marcellino [2011] but not on the factor dynamics (this alternative is explored by Camacho, Perez-Quiros and Poncela [2011, 2012]).[5]

The MS-FaMIDAS model is helpful for the short-run analysis of business cycle fluctuations. It gives both quantitative information (the GDP growth rate) and qualitative information (the state of the economy). The MIDAS specification makes it possible to incorporate within quarter information to update directly the GDP forecast and the probability of recession several times during the quarter. Moreover, this approach is implementable when some observations are missing at the end of the sample due to the publication lags. It can also deal with an irregular pattern of the missing observations (the so-called *ragged edge* problem) due to the different time releases of the indicators.

We use Monte Carlo experiments to assess the MS-FaMIDAS model relative to several benchmarks, both in-sample and out-of-sample. A particular attention is devoted to the loss due to omitting the regime switches and/or the mixed frequency data. We also compare the MS-FaMIDAS model based on distributed lag polynomials to the unconstrained-MIDAS model, as done in simple MIDAS models by Foroni et al. [2011] and in FaMIDAS models in Marcellino and Schumacher [2010]. This evaluation is made for various sets of parameters. In the out-of-sample evaluation, we use unbalanced data-sets to take into account the uneven time releases of the short term indicators. Forecasting is also performed using direct and iterative methods and results of the two approaches are compared.

We find that the new specification tracks the dynamics of the process quite accurately and captures the regime switches successfully. In contrast, there is a loss in the specifications which omits the switches of parameters or which time-aggregates higher frequency data to match the sampling rate of the lower frequency dependent variable. The unconstrained MS-FaMIDAS model is a serious competitor despite the proliferation of parameters when the lag increases. This last result is consistent with the findings of Marcellino and Schumacher [2010] and Foroni et al. [2011] and may be due to the low

[5]See also Diebold and Rudebusch [1996], Kim and Yoo [1995], Kim and Nelson [1998] and Chauvet [1998].

difference in data frequencies in our paper as in typical macroeconomic applications.

We apply the MS-FaMIDAS to model the link between the US GDP and financial variables sampled at a higher frequency. By doing that, we extend the empirical application in Guérin and Marcellino [2011]. They use the MS-MIDAS specification to assess the predictive power for US GDP growth of three financial variables taken separately: the yield curve, the S&P500 index and the Federal Funds. In this paper, we use the block of financial variables considered in Stock and Watson [2005]. The information set consists of money and credit quantity aggregates, stock prices, interest rates and spreads, exchange rates and price indexes. A real-time evaluation shows that the model with factors extracted from this dataset detects properly the US recessions at horizons up to two quarters. However, our financial factors do not help to predict quantitatively the US GDP in the short run.

The remainder of this paper is organized as follows. In section 1, we present the MS-factor MIDAS specification and describe the estimation and forecasting techniques. In the second section, we use Monte Carlo simulations to assess the in-sample and out-of-sample performance of the specification. Section 3 is devoted to the empirical application to US data. The last section offers some concluding remarks.

1 A MS-MIDAS Factor model

1.1 Specification

This section presents the MS Factor MIDAS model. We follow the notations of Clements and Galvao [2008] and Clements and Galvao [2009]. The time index t denotes the time unit of the lower frequency variable Y (the quarter in our application). We model the link between Y and higher frequency indicators X sampled m times between two time units of Y, e.g. t and $t-1$ ($m=3$ for monthly indicators as in our application). The lag operator $L^{1/m}$ operates at the higher frequency, e.g. $L^{s/m}x_t^{(m)} = x_{t-s/m}^{(m)}$.

Consider a vector of N stationary monthly series $X_t^{(m)} = \left(X_{1t}^{(m)}, X_{2t}^{(m)}, \ldots, X_{Nt}^{(m)}\right)'$, $t = 1/m, 2/m, \ldots, T$ previously standardized to mean zero and variance one. We assume that the observed variables $X_t^{(m)}$ can be decomposed into the sum of two unobserved orthogonal components: a small number of latent variables, the common factors f_t summarizing the dynamics common to all the series, and an idiosyncratic component ε_t, specific to each series. In addition, the factors can be autocorrelated. Formally, the dynamic factor is given by:

$$X_t^{(m)} = \Lambda f_t^{(m)} + \varepsilon_t^{(m)} \qquad \varepsilon_t^{(m)} i.i.d.N(0, \Sigma_\varepsilon) \tag{1}$$

$$f_t^{(m)} = A_1 f_{t-1/m}^{(m)} + \ldots + A_p f_{t-p/m}^{(m)} + B u_t^{(m)} \qquad u_t^{(m)} i.i.d.N(0, I_q) \tag{2}$$

for $t = 1/m, 2/m, \ldots, T$. In equation (1), $f_t^{(m)} = (f_{1t}^{(m)}, \ldots, f_{rt}^{(m)})'$ is a $(r \times 1)$ stationary process, Λ is an $(N \times r)$ matrix of factor loadings, $\varepsilon_t^{(m)} = (\varepsilon_{1t}^{(m)}, \ldots, \varepsilon_{Nt}^{(m)})'$ is a $N \times 1$ stationary process, (F_t) and (ε_t) are independent processes. In equation (2), the VAR process of f_t is driven by a q-dimensional standardized white noise $u_t^{(m)}$ (the dynamic shocks), $A_1, \ldots, A_p$ are $(r \times r)$ matrices of parameters and B is an $(r \times q)$ matrix.

The system of equations (1)-(2) can be cast in a state space representation. The measurement equation (1) describes the relationship between the observed variable $X_t^{(m)}$ and the unobserved state variable $f_t^{(m)}$. The state equation (2) describes how the hidden variables are generated from their lags and from innovations.

The information summarized in the latent factors is then used to forecast the lower frequency variable y_t. To relate the variable y_t to the higher frequency factors, Marcellino and Schumacher [2010] introduce the Factor MIDAS model given by:

$$y_t = \beta_0 + \beta_1 B(L^{1/m}, \theta) f_t^{(m)} + \eta_t \qquad t = 1, \ldots, T \tag{3}$$

where $f_t^{(m)}$ is a latent factor. The superscript (m) indicates that this variable is sampled at a higher frequency.

The polynomial $B(L^{1/m}, \theta)$ is the exponential Almon lag[6] with:

$$B(L^{1/m}, \theta) = \sum_{j=1}^{K} b(j, \theta) L^{(j-1)/m}, b(j, \theta) = \frac{exp(\theta_1 j + \theta_2 j^2)}{\sum_{j=1}^{K} exp(\theta_1 j + \theta_2 j^2)} \tag{4}$$

This function implies that the weights are positive. It allows a parsimonious specification since only two coefficients are needed for the K lags. The coefficient β_1 gives the impact of the factor on the dependent variable, the coefficient $\theta = \{\theta_1, \theta_2\}$ defines the lag structure. In the particular case where $\theta = \{0, 0\}$, we obtain the standard equal weighting aggregation scheme.

For r factors, the specification is given by:

$$y_t = \beta_0 + \sum_{i=1}^{r} \beta_{1,i} B(L^{1/m}, \theta_i) f_{i,t}^{(m)} + \varepsilon_t \qquad t = 1, \ldots, T \tag{5}$$

[6] Other possible specifications of the MIDAS polynomials are based on beta or step functions. See Ghysels et al. [2007] for a presentation of the various parameterizations of $B(L^{1/m}, \theta)$.

Note that the lag structure can be different for each factor. This is particularly relevant for GDP forecasting. It is possible to give a larger weight to lagged values of a factor extracted from leading indicators like financial data while more weight will be attached to the most recent values of a factor extracted from coincident indicators such as survey data or GDP components.

Like Guérin and Marcellino [2011], we extend the specification of y_t by allowing a change in the parameters of the model. We assume that the parameters of equation (5) depend on an unobservable discrete variable S_t :

$$y_t = \beta_0(S_t) + \sum_{i=1}^{r} \beta_{1,i}(S_t) B(L^{1/m}, \theta_i) f_{i,t}^{(m)} + \varepsilon_t(S_t) \qquad t = 1, \ldots, T \tag{6}$$

where $\varepsilon_t | S_t \rightarrow NID(0, \sigma^2(S_t))$. Note that the lag structure $B(L^{1/m}, \theta)$ is not regime dependent. The variable $S_t = 1, 2, \ldots, M$ represents the state that the process is in at time t. This variable is assumed to follow a first-order Markov chain defined by the following transition probabilities:

$$p_{ij} = P(S_t = i | S_{t-1} = i) \tag{7}$$

where $\sum_{j=1}^{M} p_{ij} = 1, \forall i, j = 1, 2, \ldots, M$. In the following, we only consider the case of two regimes $M = 2$.

Note that it is also possible to deal with the mixed frequencies in the state space representation of the factor model as done in Banbura and Runstler [2011]. Several papers discuss the connection between the two approaches. From a theoretical point of view, Bai et al. [2009] show that in some cases, the MIDAS representation is an exact representation of the state space approach and in other cases, it involves approximation errors that are typically small. The empirical comparison of the two approaches in Marcellino and Schumacher [2010] and Kuzin et al. [2011] shows that the MIDAS approach, more parsimonious and less prone to specification errors, performs quite well. In this paper, we do not use the integrated state space approach which appears more complicated with regime-switching parameters.

1.2 Estimation

The estimation of the MS-FaMIDAS model consists of two main steps. First, we estimate the factors. At this level, we use a method that copes with unbalanced dataset due to potential different publication lags of the higher frequency indicators. Then, we estimate the relationship between the low frequency variable and the high frequency factors.

1. Estimation of the factors (equations 1-2): we use the two-step method proposed by Doz et al. [2011] to estimate the factors in the monthly frequency. Factors are first estimated by principal components on the balanced sub-sample, i.e. over the period when all the variables X_t are known. The factors are then estimated over the entire range of observations including the period when some variables have missing observations. At this stage, we apply the Kalman filter and smoother to the state space representation. To accommodate the missing observations at the end of the sample due to publication lags, the variance of the idiosyncratic noise related to the missing observations is set to infinity (this is equivalent to skipping these observations).

2. Estimation of the MS-model (equations 6-7): we follow Guérin and Marcellino [2011] and estimate equations (6)-(7) via maximum likelihood. The likelihood is derived in the filter of Hamilton and the simplex search method is applied to find the vector of parameters maximizing the function (we use the function *fminsearch* of the Matlab's optimization toolbox). A smoothing algorithm is then applied to get a better estimation of the states. In the estimation procedure, the parameter θ_2 is constrained to be negative which guarantees a declining weight with K (see for instance Ghysels et al. [2007] for a further discussion of this issue).

1.3 Forecast

Once the specification estimated, it can be used to derive a forecast of y_t.

We consider two alternative approaches, known in the forecasting literature as the iterative and direct approaches.[7]

In the iterative approach, we exploit the dynamic structure of the factor model. The monthly factor $f_t^{(m)}$ is forecast over the quarterly horizon h (that is on hm monthly periods) with the VAR on the factor in equation (2). The forecast of y_t is then derived from an equation relating y_t to the contemporaneous values of the factors and their lags:

$$y_{t+h} = \beta_0(S_{t+h}) + \sum_{i=1}^{r} \beta_{1,i}(S_{t+h})B(L^{1/m}, \theta_i)f_{i,t+h}^{(m)} + \varepsilon_{t+h}(S_{t+h}), \qquad t = 1, \ldots, T-h \quad (8)$$

In the direct approach, no forecast of the factor is made. Instead, the forecast model of y_t is specified and estimated as a linear projection of the h-step ahead variable y_t on an intercept and the estimated factors:

$$y_{t+h} = \delta_0(S_{t+h}) + \sum_{i=1}^{r} \delta_{1,i}(S_{t+h})B(L^{1/m}, \theta_i)f_{i,t}^{(m)} + \varepsilon_{t+h}(S_{t+h}), \qquad t = 1, \ldots, T-h \quad (9)$$

[7] See Chevillon and Hendry [2005] and Marcellino, Stock and Watson [2006] for a recent discussion on this issue in single-frequency models.

The two equations are estimated for $t, \ldots, T-h$ and the forecast of y_t in $T+h$ is derived by weighting each estimated regime with the predicted probabilities of the two states in $T+h$. The forecast of the chain S_t at horizon h is given by:

$$P(S_{T+h} = 1 \mid I_T; \Theta) = (p_{11} + p_{22} - 1)^h (P(S_T = 1 \mid I_T; \Theta) - \xi_1) + \xi_1 \qquad (10)$$

where the last term is the unconditional probability of state 1 given by $\xi_1 = \frac{1-p_{22}}{2-p_{11}-p_{22}}$.

The MS-FaMIDAS model provides a useful tool in the context of short-run GDP forecasting and detection of recession. First, the MIDAS regression incorporates indicators sampled m times during the basic time unit. Hence, the MIDAS specification makes it possible to incorporate within quarter information and to update the GDP forecast and the probability of the state m times during the quarter in a very direct way. Moreover, this approach is implementable when some observations are missing at the end of the sample due to the publication lags through the application of the Kalman filter. It can also deal with an irregular pattern in the missing observations (the so-called *ragged edge* problem) due to the different time releases of the indicators. More generally, the Kalman filter also allows us to exploit the information provided by variables available on different sample periods.

2 Monte Carlo simulations

2.1 In-sample evaluation

This section presents the results of the in-sample evaluation of the model. At this stage, the specification is estimated on the whole sample and we use a balanced dataset.

Our Monte Carlo experiment involves the following steps.

1. Simulations of a MS-factor MIDAS model:

a. Simulation of r-dimensional factors $f_t^{(m)}$, $t = 1/m, 2/m, \ldots, T$ following a VAR(p) dynamics in which errors are generated via a pseudo-random number generator and distributed $N(0,1)$.

b. Construction of N observable variables $x_{it}^{(m)}$ according to $x_{it}^{(m)} = \lambda_i' f_t^{(m)} + e_{it}^{(m)}$, $t = 1/m, 2/m, \ldots, T$ where λ_i and $e_{it}^{(m)}$ are assumed *i.i.d.* normal.

c. Simulation of the low frequency variable y_t, $t = 1, 2, \ldots, T$ according to equation (6) where S_t is a simulated first-order Markov chain.

2. Estimation of the relationship between the low and high frequency variables $y_t = g(f_t^{(m)}) + \varepsilon_t$, $t = 1, \ldots, T$ with alternative specifications of $g(.)$ detailed below.

We simulate T observations of the low frequency variable Y and $T \times m$ observations of the low frequency indicators X, with m the number of times the high frequency indicators

X are sampled between two time units of Y. We replicate these steps $R = 1000$ times. Note that in steps 1a and 1c, the first 100 simulated observations of the factors f_t and the Markov chain S_t are discarded to remove the effect of the initial conditions.

Several specifications are estimated from the simulated observations of y_t. First, we estimate the MS-factor MIDAS model in order to assess the robustness of the estimation procedure. We also consider alternative specifications to measure the loss due to information aggregation and/or omission of the non-linear dynamics. Formally, six models are considered. The first three specifications are linear and the last three equations are MS models (the last one is the MS-FaMIDAS specification which is the true model):

$$y_t = \beta_0 + \sum_{i=1}^{r} \sum_{j=1}^{\lceil K/m \rceil} \beta_{i,j} L^j \hat{f}_{i,t} + \varepsilon_t \qquad (ML1)$$

$$y_t = \beta_0(S_t) + \sum_{i=1}^{r} \sum_{j=1}^{\lceil K/m \rceil} \beta_{i,j}(S_t) L^j \hat{f}_{i,t} + \varepsilon_t(S_t) \qquad \text{(MS1)}$$

$$y_t = \beta_0 + \sum_{i=1}^{r} \sum_{j=1}^{K} \beta_{i,j} L^{j/m} \hat{f}_{i,t}^{(m)} + \varepsilon_t \qquad (ML2)$$

$$y_t = \beta_0(S_t) + \sum_{i=1}^{r} \sum_{j=1}^{K} \beta_{i,j}(S_t) L^{j/m} \hat{f}_{i,t}^{(m)} + \varepsilon_t(S_t) \qquad \text{(MS2)}$$

$$y_t = \beta_0 + \sum_{i=1}^{r} \beta_{1,i} B(L^{1/m}, \theta_i) \hat{f}_{i,t}^{(m)} + \varepsilon_t \qquad (ML3)$$

$$y_t = \beta_0(S_t) + \sum_{i=1}^{r} \beta_{1,i}(S_t) B(L^{1/m}, \theta_i) \hat{f}_{i,t}^{(m)} + \varepsilon_t(S_t) \qquad \text{(MS3)}$$

where the polynomial $B(L^{1/m}, \theta)$ is defined in equation (4).

In equations (ML1) and (MS1), the factors are converted to quarterly frequency by averaging the months of the quarter. We choose a number of quarterly lags consistent with the true monthly lag, given by the closest quarterly lag larger than or equal to the monthly lag in the DGP. The comparison of these two equations with the following ones allows measuring the loss due to information aggregation. The equations (ML2) and (MS2) are MIDAS models with unrestricted lag polynomials also considered in Marcellino and Schumacher [2010] (and initially proposed by Koenig, Dolmas and Piger [2003]).[8] In equations (ML3) and (MS3), we use the Almon polynomial as defined in equation (4) to get a more parsimonious specification. The specifications (ML2) and (MS2) do not impose any structure on the coefficients of the lagged factors as the one implied by the exponential Almon function but are far less parsimonious. For instance, for $K = 12$ and $r = 1$, we need to estimate 30 parameters in the MS unconstrained model (MS2) against only 10 parameters in the MS Almon specification (MS3).

[8]Foroni et al. [2011] also compare the MIDAS specification based on distributed lag polynomials to the unconstrained-MIDAS model with a single high frequency indicator. They study the relative performance of the two specifications on simulated data and for nowcasting euro area and US GDP. They show that the MIDAS model with unrestricted lag polynomials can outperform the restricted MIDAS model especially for small differences in sampling frequencies (i.e. for small m).

We use different sets of parameters. The reference one is chosen close to the empirical setup in section 3 with $m = 3$, a sample size $T = 200$ quarters (i.e. 600 observations for the high frequency variables for $m = 3$), $r = 1$ factor driven by $q = 1$ shock and extracted from $N = 50$ monthly variables. We assume that the factor follows an AR(1) process where the autoregressive coefficient φ is equal to -0.3 (such a value is relevant for a factor extracted from financial data). The coefficients of the MS-MIDAS factor model used in simulating the dependent variable are given below:

$(p_{11}, p_{22}, \beta_{0,1}, \beta_{1,1}, \beta_{0,2}, \beta_{1,2}, \theta_1, \theta_2, \sigma_1, \sigma_2) = (0.95, 0.85, 0.5, -1, -0.5, 1, 2, -0.15, 0.3, 0.2)$

In our application to the US output growth rate, the shorter state 2 characterized by a lower mean and a lower volatility corresponds to the recession regime.

In Table 1, we assess the quality of the estimation when the DGP is correctly identified. To this end, we report the average estimates of the coefficients of the MS-factor MIDAS and the standard deviations of the estimates in the 1000 replications. For the parameters θ_1 and θ_2, we also report an average measure of the error on the weights given by:

$$\frac{\sum_{j=1}^{K}[b(j,\hat{\theta}) - b(j,\theta)]^2}{\sum_{j=1}^{K} b(j,\theta)^2} \tag{13}$$

As noted by Guérin and Marcellino [2011], it is more important to correctly estimate the shape of the function rather than the point estimates of θ_1 and θ_2.

When we choose the true specification, our estimation procedure provides accurate estimates of the parameters. Indeed, the average estimates are generally very close to the parameters of the underlying DGP and the dispersion is low. Note that the volatility of the estimated parameters of the shortest regime is larger. This is not surprising since this regime is less frequently visited. The estimated parameters of the Almon function, θ_1 and θ_2, are also less accurate, especially for small values of K and the dispersion of the estimates of these two parameters is higher. However, the approximate error remains very low even for the smallest values of K. In addition, the quality of adjustment is very high as shown by the high values of the R-squared. This quality decreases with the Almon lag K which can be related to the increase in the approximation error for large K.

In Table 2, we assess the consequence of changes in the reference setup on the estimation accuracy and the relative performance of the six specifications. We consider alternatively a lower persistence of the recession regime $p_{22} = 0.70$, a smaller sample size $T = 120$ quarters (i.e. 360 months), different numbers of variables, $N = 25$ and $N = 100$, a flatter weighting function obtained for smaller values of $\theta = \{0.2, -0.015\}$ and $r = 2$ factors driven by $q = 2$ shocks. We also compare the less persistent AR(1) process for the factor with $\varphi = -0.3$ (suitable for factors extracted from financial data) to a more persistent one ($\varphi = 0.8$ more appropriate for real and survey data). Finally, we consider a larger difference in sampling frequencies $m = 12$ (which corresponds to the case of higher frequency data sampled at weekly frequency when the lower frequency variable is available on a quarterly basis).[9] We report the results of all these configurations for $K = 3, \ldots, 12$.

We apply several criteria in order to compare the ability of the six specifications to capture the dynamics of y_t and S_t. First, we use the traditional R-squared and Bayesian information criteria. In the case of MS models, the R-squared is derived by weighting the residuals of each regime with the predicted probability $P(S_t \mid I_{t-1}; \Theta)$. To assess the quality of regime estimation, we also use the quadratic probability score (QPS) given by:

$$\frac{1}{T} \sum_{t=1}^{T} (P(S_t = 1 \mid I_T; \Theta) - S_t)^2 \tag{14}$$

with $P(S_t = 1 \mid I_T; \Theta)$ the smoothed probability of state one.

In order to assess the regime estimation, we will also consider a new criterion: a Turning Point Indicator (TPI hereafter). This indicator aims at evaluating the ability of the model to detect each turning point accurately or with a lead / lag of τ quarter.

$$TPI(\lambda, \tau) = \frac{1}{n} \sum_{t=1}^{T} \max_{-\tau \leq h \leq \tau} [(P_{t-h}(\lambda) - P_{t-h-1}(\lambda))(S_t - S_{t-1})] \tag{15}$$

where n is the number of observed turning points, $P_t(\lambda) = (P(S_t = 1 \mid I_T; \Theta) > \lambda)$ with the threshold parameter λ taken equal to 0.5 or 0.4 in our application. Compared to the QPS criterion, this index focuses on the periods with a switch of regime.

Overall, we find a loss when converting the high frequency indicators to the lower frequency with simple time averaging (ML1 relative to ML2 and ML3 and MS1 relative to MS2 and MS3). This loss is larger for small K. The improvement of the R-squared can be up to 20% with the MIDAS specification. The two regimes are also better identified in the MIDAS specification as indicated by the lower values of the QPS criterion. There is also a marked loss when ignoring the non-linear dynamics (MLi relative to MSi). The decrease of the R-squared is up to 90% in MLi relative to MSi (or 65% for the BIC criterion penalizing the number of parameters). Finally, we find that the two MIDAS specifications perform similarly in terms of quality of adjustment and regime identification, although MS3 is much more parsimonious than MS2. This is less true when we consider a larger difference between frequencies $m = 12$, as already noted by Foroni et al. [2011] in a linear framework and with a single explanatory variable.

[9] In this particular case, we consider a higher number of lags K for X in order to condition Y on the same number of quarterly lags of X and we simulate $T \times m = 2400$ weekly observations of X.

Table 1: Evaluation of the estimation with Monte Carlo simulations

K	$\phi = -0.3$	$p_{11} = 0.95$	$p_{22} = 0.85$	$\beta_{0,1} = 0.5$	$\beta_{1,1} = -1$	$\beta_{0,2} = -0.5$	$\beta_{1,2} = 1$	$\theta_1 = 2$	$\theta_2 = -0.15$	$\sigma_1 = 0.3$	$\sigma_2 = 0.2$	apx err	R^2
3	-0.292	0.948	0.829	0.499	-0.987	-0.501	0.987	4.618	-0.672	0.316	0.219	0.002	0.89
	(0.039)	(0.021)	(0.076)	(0.028)	(0.053)	(0.039)	(0.068)	(12.323)	(2.463)	(0.02)	(0.031)		
4	-0.292	0.949	0.832	0.501	-0.990	-0.500	0.990	2.516	-0.223	0.313	0.215	0.003	0.88
	(0.039)	(0.019)	(0.064)	(0.029)	(0.060)	(0.039)	(0.074)	(4.133)	(0.59)	(0.019)	(0.028)		
5	-0.289	0.949	0.829	0.500	-0.993	-0.496	0.994	2.135	-0.165	0.309	0.209	0.004	0.86
	(0.038)	(0.021)	(0.075)	(0.028)	(0.066)	(0.092)	(0.117)	(0.92)	(0.105)	(0.019)	(0.028)		
6	-0.291	0.949	0.829	0.501	-0.991	-0.500	0.989	2.102	-0.159	0.307	0.207	0.006	0.83
	(0.039)	(0.019)	(0.073)	(0.029)	(0.072)	(0.041)	(0.098)	(0.693)	(0.067)	(0.019)	(0.029)		
7	-0.293	0.949	0.827	0.499	-0.993	-0.502	0.991	2.067	-0.155	0.304	0.202	0.007	0.81
	(0.038)	(0.020)	(0.078)	(0.028)	(0.077)	(0.040)	(0.109)	(0.536)	(0.045)	(0.019)	(0.029)		
8	-0.293	0.948	0.825	0.501	-0.991	-0.502	0.991	2.064	-0.155	0.301	0.201	0.011	0.80
	(0.040)	(0.021)	(0.074)	(0.028)	(0.085)	(0.040)	(0.125)	(0.453)	(0.035)	(0.018)	(0.028)		
9	-0.290	0.949	0.828	0.499	-0.990	-0.499	0.997	2.046	-0.153	0.302	0.199	0.010	0.79
	(0.038)	(0.02)	(0.071)	(0.029)	(0.093)	(0.043)	(0.141)	(0.42)	(0.032)	(0.02)	(0.029)		
10	-0.292	0.949	0.831	0.501	-0.988	-0.499	0.991	2.056	-0.154	0.301	0.200	0.015	0.78
	(0.039)	(0.019)	(0.071)	(0.028)	(0.102)	(0.041)	(0.145)	(0.421)	(0.031)	(0.019)	(0.027)		
11	-0.293	0.949	0.829	0.500	-0.992	-0.500	0.991	2.043	-0.153	0.300	0.199	0.012	0.78
	(0.038)	(0.019)	(0.074)	(0.027)	(0.101)	(0.046)	(0.151)	(0.410)	(0.030)	(0.019)	(0.026)		
12	-0.290	0.949	0.828	0.501	-0.990	-0.502	0.995	2.057	-0.154	0.301	0.198	0.015	0.78
	(0.039)	(0.020)	(0.069)	(0.028)	(0.108)	(0.043)	(0.152)	(0.402)	(0.030)	(0.019)	(0.027)		

Note: This table presents the average estimate and in brackets the standard deviation of the estimates. The last two columns report the approximation error of the weights and the R-squared.

Table 2: Relative performance of the six specifications in Monte Carlo simulations

		R^2						BIC						QPS			apx err	
DGP	K	ML1	ML2	ML3	MS1	MS2	MS3	ML1	ML2	ML3	MS1	MS2	MS3	MS1	MS2	MS3	ML3	MS3
benchmark	3	0.10	0.21	0.21	0.60	0.89	0.89	498.02	479.33	479.93	443.46	214.61	206.42	0.057	0.015	0.015	0.072	0.002
	6	0.12	0.17	0.15	0.77	0.84	0.83	403.69	413.95	401.08	264.68	224.96	190.41	0.023	0.017	0.016	0.140	0.006
	9	0.09	0.13	0.10	0.75	0.79	0.79	356.66	378.95	353.09	226.25	242.59	182.54	0.020	0.019	0.017	0.335	0.010
	12	0.08	0.14	0.10	0.75	0.79	0.78	357.80	388.48	349.59	236.13	270.17	181.95	0.020	0.020	0.017	0.396	0.015
$p_{22} = 0.70$	3	0.19	0.39	0.39	0.52	0.88	0.88	462.53	412.62	413.05	438.11	223.07	216.07	0.074	0.019	0.018	0.019	0.002
	6	0.23	0.30	0.28	0.72	0.81	0.80	353.02	355.29	342.46	268.14	231.60	199.87	0.028	0.023	0.02	0.039	0.007
	9	0.16	0.21	0.18	0.69	0.74	0.73	308.08	326.67	301.12	234.26	248.03	194.88	0.025	0.025	0.020	0.092	0.012
	12	0.15	0.21	0.17	0.68	0.73	0.72	309.22	336.38	297.56	241.04	271.12	191.73	0.025	0.026	0.020	0.132	0.014
$T = 120$	3	0.12	0.24	0.23	0.59	0.89	0.89	298.68	286.75	287.41	273.11	139.21	134.21	0.072	0.016	0.016	0.122	0.004
	6	0.14	0.20	0.18	0.76	0.84	0.83	241.08	250.35	239.13	165.81	147.77	123.43	0.025	0.019	0.017	0.280	0.010
	9	0.10	0.17	0.13	0.74	0.79	0.78	214.57	234.42	211.32	145.99	159.50	120.50	0.023	0.023	0.018	0.554	0.018
	12	0.10	0.18	0.12	0.74	0.79	0.78	216.45	244.02	209.37	154.74	181.71	118.55	0.024	0.024	0.018	0.605	0.026
$N = 25$	3	0.10	0.20	0.20	0.59	0.88	0.88	499.56	483.05	483.63	449.04	238.86	230.85	0.056	0.016	0.016	0.082	0.003
	6	0.12	0.16	0.15	0.76	0.83	0.82	404.92	415.38	402.60	272.34	237.52	202.85	0.024	0.019	0.018	0.185	0.007
	9	0.08	0.12	0.10	0.75	0.79	0.78	359.23	381.63	355.69	231.65	249.90	190.45	0.021	0.020	0.017	0.363	0.014
	12	0.08	0.13	0.09	0.74	0.79	0.78	358.86	389.81	350.79	241.01	276.71	189.07	0.021	0.020	0.017	0.402	0.016
$N = 100$	3	0.10	0.21	0.21	0.60	0.90	0.90	497.93	479.43	480.08	443.00	199.22	191.07	0.056	0.014	0.014	0.080	0.002
	6	0.12	0.17	0.15	0.77	0.85	0.84	404.45	414.80	401.97	260.35	215.61	181.35	0.023	0.017	0.016	0.179	0.006
	9	0.08	0.12	0.10	0.76	0.80	0.79	359.92	382.59	356.42	222.12	237.26	177.88	0.020	0.018	0.016	0.367	0.010
	12	0.09	0.14	0.10	0.75	0.79	0.78	355.68	386.78	347.44	232.26	265.92	177.86	0.020	0.019	0.016	0.422	0.012
$\theta_1 = 0.2$	3	0.14	0.16	0.15	0.83	0.84	0.84	396.52	404.33	404.67	191.93	199.33	191.13	0.016	0.016	0.016	0.135	0.002
$\theta_2 = -0.015$	6	0.10	0.12	0.10	0.79	0.80	0.79	352.55	369.40	356.62	185.67	215.56	182.71	0.017	0.018	0.017	0.330	0.008
	9	0.07	0.11	0.08	0.77	0.79	0.77	342.77	367.95	342.05	190.35	236.89	179.48	0.017	0.019	0.016	0.591	0.014
	12	0.06	0.10	0.06	0.76	0.78	0.76	339.19	373.07	333.61	199.90	260.75	179.20	0.017	0.020	0.016	0.847	0.016
$r = 2$	3	0.21	0.25	0.24	0.81	0.90	0.90	534.51	543.49	547.17	346.82	273.99	262.09	0.034	0.016	0.015	0.350	0.202
	6	0.21	0.26	0.22	0.84	0.87	0.87	527.67	557.14	536.12	330.01	378.51	287.29	0.025	0.024	0.020	0.552	0.151
	9	0.23	0.29	0.23	0.81	0.86	0.87	529.47	576.31	528.33	363.01	441.92	280.39	0.028	0.025	0.019	0.546	0.127
	12	0.23	0.31	0.22	0.81	0.85	0.87	535.98	599.53	526.67	388.44	509.84	280.72	0.031	0.027	0.021	0.527	0.125
$\rho = 0.8$	3	0.23	0.26	0.25	0.83	0.93	0.93	549.79	550.85	551.67	341.02	215.98	208.02	0.039	0.013	0.012	0.189	0.005
	6	0.25	0.27	0.26	0.91	0.93	0.93	528.64	543.98	531.44	224.24	229.05	194.09	0.016	0.013	0.013	0.355	0.009
	9	0.24	0.26	0.24	0.91	0.92	0.92	513.84	538.49	512.63	219.23	248.16	186.14	0.017	0.015	0.014	0.525	0.008
	12	0.23	0.27	0.24	0.91	0.92	0.92	514.30	547.24	508.30	223.96	271.80	182.10	0.018	0.016	0.014	0.648	0.006
$m = 12$	12	0.04	0.14	0.10	0.69	0.79	0.78	351.01	388.99	350.08	257.85	271.67	183.41	0.020	0.016	0.014	0.471	0.016
	24	0.05	0.19	0.10	0.69	0.79	0.78	355.23	440.06	350.00	268.66	398.94	183.20	0.020	0.016	0.014	0.464	0.014
	36	0.05	0.25	0.10	0.69	0.79	0.78	359.47	491.19	350.03	278.93	528.34	183.47	0.020	0.016	0.014	0.439	0.016
	48	0.06	0.31	0.10	0.69	0.79	0.78	363.82	542.53	350.10	289.82	658.84	183.47	0.020	0.015	0.014	0.436	0.013

In the following, we further assess the relative performance of the six specifications and the impact of the parameters by focusing on the BIC criterion of the last five specifications relative to the linear model (ML1).

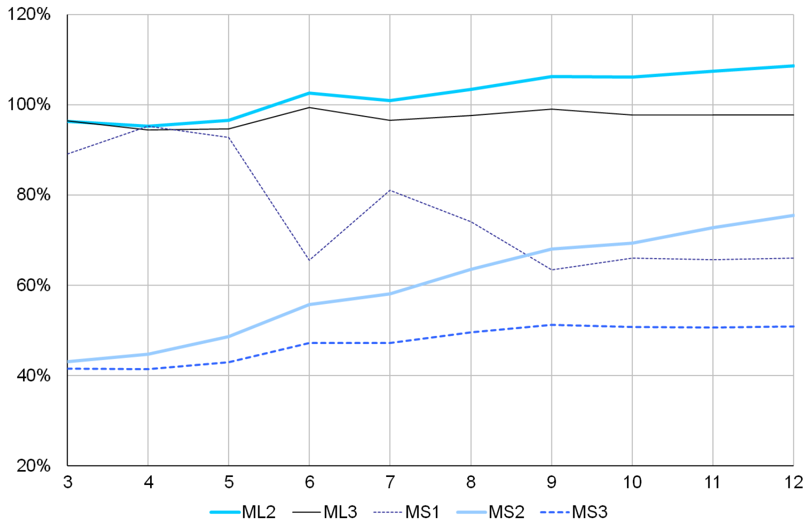


Figure 1: BIC relative to ML1 for the benchmark setup

Figure 1 depicts the BIC ratio against the simple linear model (ML1) for the benchmark parameters and $K = 3, \dots, 12$. No significant difference between the linear specifications is found, even if we note a slight advantage of the parsimonious Almon MIDAS (ML3) for large K. Hence, the loss due to aggregating information may not be apparent when ignoring the switches of the parameters. The unconstrained specification is performing even worse for large values of K. When comparing the non-linear specifications, the MS1 model shows a relative poor performance for small values of the order K. This can be due to the unevenly distributed weights of the lags in the DGP. This pattern leads to a large loss when the information is simply time-averaged. This loss is particularly large for K non-multiple of 3, since MS1 gives an equal weight to each monthly lag of the quarters including the last one(s) not present in the DGP. Nevertheless, the relative parsimony of MS1 compared to the unconstrained specification MS2 gives an advantage to the former one for large values of K. For instance, for $K = 12$, even though MS2 still outperforms ML1, it is heavily penalized by the extra 16 parameters. Finally, MS3 outperforms the other specifications by 23% to 53% for the largest value of K. This is not surprising since the data generating process is a MS-factor MIDAS. However the difference with the unconstrained model is low for small values of K due to the flexibility and relative parsimony of MS2.[10]

[10] Foroni et al. [2011] find similar results in a simpler framework.

In Figure 2, we check whether the relative advantage of our model is robust to the choice of the DGP. The relative performance of MS3 is still measured by the BIC ratio of this model against the linear specification (ML1). Some changes in the parameters have an impact on the performance of MS3. First, the sample size effect seems to be larger in the non-linear model than for the estimation of the linear benchmark. Indeed, the relative performance of the former model diminishes by roughly 5 pp for all K when the sample size is reduced from 50 to 40 years. Second, the simulations show that the linear approximation is less detrimental when the volatile regime is shorter-lived (with $p_{22} = 0.70$, the expected duration of state 2 is of 3 quarters compared to a rough 7 quarters for $p_{22} = 0.85$). The relative gain as measured by the BIC ratio loses 12 percentage points for $K = 12$ to end up at 62%. The impact of the persistence in the factor dynamics is larger. The simulation results show an increasing enhancement of the MS-MIDAS estimation accuracy relative to the linear model: the gain rises from 4pp for $K = 3$ to 15pp for $K = 12$. This result is probably due to a clearer and accordingly better estimation of the non-linear dynamics with a more persistent factor.

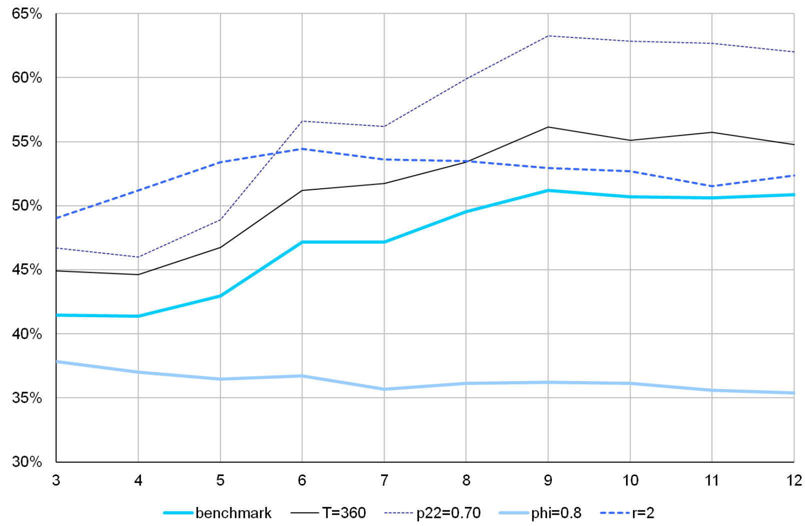
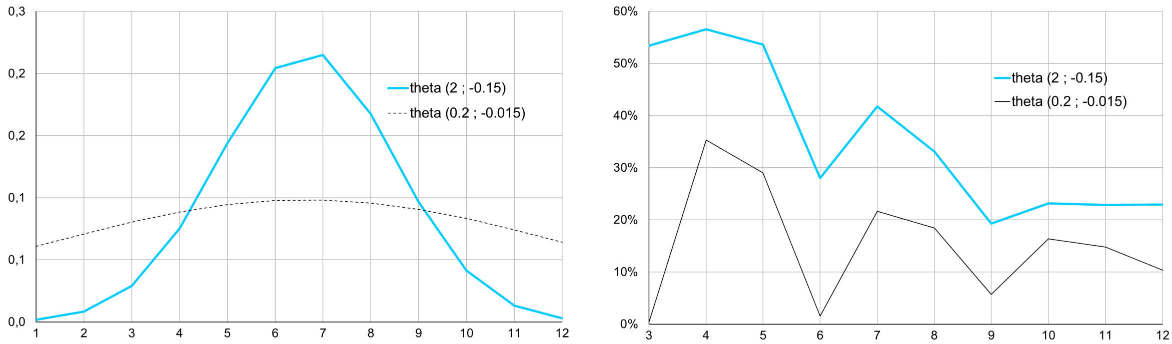


Figure 2: BIC relative to ML1 for various setups

In Figure 3, we focus on the impact of the shape of the exponential Almon function. We compare the results for two MS-FaMIDAS DGPs with $\theta = \{2, -0.15\}$ and $\theta = \{0.2, -0.015\}$. In the left-hand graph, we plot the two alternative exponential Almon functions and the BIC criterion relative to MS1 for the two specifications in the right-hand graph. As expected, choosing a flatter distribution of the weights leads to a smaller gain of the MIDAS approach compared to the trivial aggregation taking the mean over the quarter. For a distribution with less variation of the weights within the quarter, no significant gain is found with the MIDAS approach for K multiple of 3. Nevertheless, the MIDAS approach performs better for DGPs where information beyond the last quarter is relevant (e.g. $K = 4$ and $K = 5$). For instance, the gain stands at 35% for $K = 4$ after 0% for $K = 3$. Indeed, the MIDAS specification excludes the irrelevant months while MS1 gives the same weight to the three months of each quarter.



(a) Almon distribution

(b) BIC relative to MS1

Figure 3: The impact of the weight distribution

Finally, Figure 4 illustrates the impact of the number of cross-sections N. The assessment of the sensitivity to the number of variables N is motivated by a recent debate in the factor literature. Traditionally, factors are extracted from a large database but Boivin and Ng [2006] and Bai and Ng [2008] show that increasing the number of variables can be detrimental to the forecast accuracy. In our simulations, the quality of adjustment tends to increase with the size of the database from which the factors are extracted. This gain decreases with the lag order K in the DGP. Thus, little impact of the size of the database may be found for large values of K.

2.2 Forecast evaluation

We now turn to the out-of-sample evaluation of the forecasts of the model at different horizons and with unbalanced data-sets.

The experimental design is the following. We still generate data from a MS-factor MIDAS model as described in the previous section but we remove the last monthly observations of the simulated sample. We estimate the six specifications (ML1)-(ML3) and (MS1)-(MS3) over the rest of the period and forecast the variable y_t at horizon h with the direct or iterative approach. We expand recursively the sample and repeat these calculations up to the last quarter of the out-of-sample period. We finally get three sets

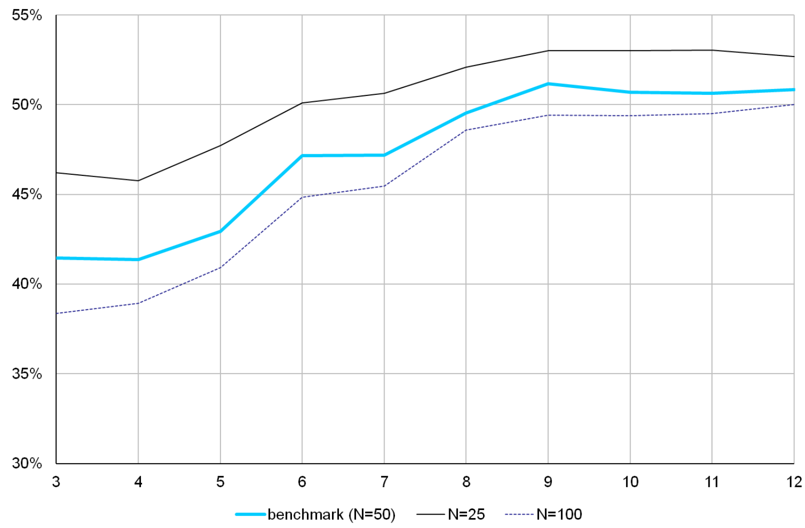


Figure 4: The impact of the database size

of forecasts at horizon h made at each month of the quarter, that is hm forecasts for each quarterly observation of y_t. We replicate R times the whole procedure for different forecast horizons h.

In real-time applications, the datasets typically contain missing observations for certain time series at the end of the period due to different publication lags. To address this issue, we do not use a balanced dataset. We suppose instead that the set of N monthly indicators is released with various delays of publication, ranging from 0 to 2 months. The delays that we consider are typical of those found in the context of short run forecasting. For this purpose, the practitioner uses survey data and financial series available during the month to which they refer, while hard indicators such as retail sales or industrial production index are released with a delay of one or two months. In the recursive scheme, we replicate the pattern of missing values at the end of each sample.

We use the reference parameters considered in the in-sample evaluation with $N = 50$ and $T = 200$. We provide the results for two alternative lags of the high frequency variable $K = 5$ (non multiple of three) and $K = 12$. In addition, we suppose that among the 50 monthly indicators, 30 are published during the reference month, 15 with a delay of one month and 5 with a delay of two months. These proportions correspond to the composition of the sample in our empirical application where we use a majority of financial variables. The out-of-sample window contains 120 monthly observations and we consider three forecast horizons $h = 0$ (nowcasting), $h = 1$ and $h = 2$ (forecasting).

Again, several criteria are applied to assess the forecast of y_t and S_t. First, we use the usual root mean squared forecast error (RMSFE) to assess the quality of the forecast of y_t. To measure the quality of regime forecast, we also use the quadratic probability score (QPS) and the turning point indicator (TPI) defined in (14) and (15) where the smoothed probability is replaced by the prediction of the Markov chain (10). In the TPI, we use a threshold parameter λ equal to 0.5 and a lag / lead $\tau = 2$.[11] Using these criteria, we compare the performance of the six models and the two usual benchmarks: an autoregressive process of order 1 with a constant and a random walk with a drift. In the case of the AR process, we use a one-period-ahead model iterated forward for the desired number of periods. The forecast derived from the random walk is obtained as the average of the past GDP growth rate computed at every recursion.

The results obtained for $R = 200$ replications are shown in Table 3. Overall, the findings of the in-sample analysis remain valid.

Among the eight specifications, MS3 provides the best quantitative forecasts (RMSE criterion). The model does not outperform the random walk for large forecast horizons but the quality of forecasts gradually increases as the horizon shortens and more information is available on the quarter to be forecast. The aggregation of information degrades the forecast (ML1 versus ML2-3 and MS1 versus MS2-3). The difference is sharper for $K = 5$ non multiple of three as found in the previous section. Similarly, the omission of the non-linearity worsens the criteria (ML relative to MS). Among the MS models, the performance of MS2 is fairly close to that of MS3 even for $K = 12$ despite the proliferation of parameters for large values of K.

Regarding the forecast of the chain (QPS and TPI criteria), the results are also supportive of the mixed frequency models (MS2 and MS3). The QPS criterion is smaller in the MS2 and MS3 specifications and the proportion of detected turning points is higher. Again, the differences are striker for $K = 5$. When $K = 12$, the QPS criteria are not different in the three models but the MS3 specification outperforms the MS1 model according to the TPI criterion for short horizons. The performance of MS2 and MS3 is also very close according to the two criteria.

At last, we do not find strong differences between the iterative and direct approaches. The direct approach provides more accurate forecasts at shorter horizons for $K = 5$ (except for $h = 0$ where the results are identical by construction). In contrast, the iterative approach performs better for large h. For $K = 12$, the results are more supportive of the direct approach. The iterative approach performs slightly better only for $h = 2$.

[11] The results are qualitatively similar for other values of these two parameters. They are not reported here for the sake of parsimony.

Table 3: Out-of-sample performance of the six specifications on simulated data

		Direct approach							Iterative approach						
h		2	5/3	4/3	1	2/3	1/3	0	2	5/3	4/3	1	2/3	1/3	0
K=5															
	RW	**0**.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75
	AR	**0**.75	**0**.74	0.74	0.74	0.73	0.73	0.73	0.75	0.74	0.74	0.74	0.73	0.73	0.73
R	ML1	0.76	0.86	0.75	0.74	0.74	0.74	0.74	0.76	0.77	0.75	0.74	0.74	0.74	0.74
M	ML2	0.77	0.76	0.72	0.70	0.69	0.70	0.69	0.75	0.75	0.71	0.69	0.70	0.70	0.69
S	ML3	0.76	0.76	0.73	0.70	0.69	0.69	0.69	0.75	0.75	0.71	0.69	0.72	0.69	0.69
E	MS1	**0**.75	0.90	0.71	0.70	0.67	0.67	0.67	0.74	0.75	0.71	0.69	0.66	0.66	0.67
	MS2	0.76	0.75	0.73	0.69	0.57	0.52	0.51	0.73	0.73	0.62	0.59	0.64	0.52	0.51
	MS3	**0**.75	**0**.74	**0**.64	**0**.60	**0**.52	**0**.52	**0**.51	**0**.73	**0**.71	**0**.62	**0**.59	**0**.59	**0**.52	**0**.51
Q	MS1	0.18	0.15	0.15	0.15	0.12	0.12	0.12	0.17	0.15	0.15	0.15	0.12	0.12	0.12
P	MS2	0.21	0.19	0.19	0.17	0.11	0.09	0.09	0.15	0.17	0.13	0.13	0.14	0.09	0.09
S	MS3	0.21	0.20	0.14	0.13	0.09	0.09	0.09	0.15	0.15	0.13	0.13	0.12	0.09	0.09
T	MS1	0.17	0.35	0.34	0.35	0.60	0.60	0.60	0.08	0.34	0.34	0.34	0.60	0.60	0.60
P	MS2	0.20	0.40	0.40	0.49	0.75	0.83	0.83	0.05	0.46	0.54	0.59	0.61	0.82	0.83
I	MS3	0.26	0.42	0.56	0.54	0.76	0.82	0.80	0.05	0.59	0.54	0.59	0.70	0.82	0.80
K=12															
	RW	0.59	0.59	0.59	0.59	0.59	0.59	0.59	0.59	0.59	0.59	0.59	0.59	0.59	0.59
	AR	0.58	0.57	0.57	0.57	0.56	0.56	0.56	0.58	0.57	0.57	0.57	0.56	0.56	0.56
R	ML1	0.59	0.60	0.58	0.58	0.58	0.58	0.58	0.59	0.58	0.58	0.58	0.58	0.58	0.58
M	ML2	0.61	0.60	0.59	0.59	0.59	0.59	0.59	0.59	0.59	0.59	0.58	0.58	0.58	0.59
S	ML3	0.59	0.58	0.58	0.58	0.57	0.57	0.57	0.59	0.58	0.58	0.57	0.58	0.58	0.57
E	MS1	**0**.56	0.55	0.53	0.52	0.46	0.45	0.45	**0**.55	**0**.52	**0**.51	0.51	**0**.45	**0**.45	0.45
	MS2	**0**.58	0.55	0.54	0.53	0.45	0.45	0.45	**0**.55	0.53	**0**.51	**0**.50	0.47	**0**.45	0.45
	MS3	**0**.56	**0**.51	**0**.50	**0**.50	**0**.43	**0**.43	**0**.43	**0**.55	0.53	0.52	**0**.50	0.46	0.46	**0**.43
Q	MS1	0.16	0.13	0.13	0.13	0.09	0.09	0.09	0.15	0.13	0.13	0.13	0.09	0.09	0.09
P	MS2	0.16	0.14	0.14	0.13	0.09	0.09	0.09	0.15	0.13	0.13	0.13	0.10	0.09	0.09
S	MS3	0.16	0.13	0.13	0.13	0.09	0.09	0.09	0.15	0.13	0.13	0.13	0.10	0.10	0.09
T	MS1	0.08	0.52	0.54	0.52	0.69	0.68	0.70	0.04	0.49	0.51	0.49	0.67	0.68	0.70
P	MS2	0.12	0.41	0.56	0.51	0.68	0.68	0.69	0.06	0.48	0.51	0.52	0.68	0.68	0.69
I	MS3	0.04	0.52	0.52	0.56	0.72	0.71	0.71	0.07	0.46	0.52	0.52	0.66	0.70	0.71

This table reports the RMSE, QPS and TPI criteria. The TPI criterion is given for $\lambda = 0.5$ and $\tau = 2$.

3 Forecasting the US GDP

3.1 The data

The database consists of the US real output growth (quarterly) and the block of financial variables (monthly) considered in Stock and Watson [2005].[12]

Our dataset was collected in September 2011. The financial variables (except NAPM) are taken from Datastream. The database includes money and credit quantity aggregates, stock prices, interest rates and spreads, exchange rates and price indexes. The 56 variables are listed in Appendix 1, together with their source and their transformation. In addition, we use vintages of output growth from the real-time datasets due to Croushore and Stark [2001] (see also Croushore [2011] for a recent survey on the use of real time data). We suppose that the financial variables are not revised. In the out-of-sample evaluation of the model, we will compare the GDP forecasts to the final values of output growth approximated by the final vintage available in September 2011.

All series have been transformed to stationarity by taking logarithm, difference, or log-difference. In addition, all series are standardized to mean zero and variance one. The sample period is 1959Q1-2010Q4 and includes eight recessions. We use the dates of recessions identified by the National Bureau of Economic Research (NBER). The financial variables are released during the month to which they refer, while the monetary aggregates and the price indices have a publication lag of one month (in the out-of-sample evaluation, we make the assumption that the delays in their publication do not change over time).

3.2 Results of estimation

We now turn to the results of estimation of the MS-factor MIDAS model on US data.

In order to capture the changes in the volatility pattern in the business cycle fluctuations over the last 50 years, we slightly modify the MS-FaMIDAS specification considered in the previous section. We introduce a double break in the model: first, we allow a decrease in the variance parameters from 1984Q1 to 2007Q3 related to the so-called 'great moderation' (Kim and Nelson [1999], McConnell and Perez-Quiros [2000]). In this paper, we also allow an increase in the volatility after 2007Q4 given the sharp increase in volatility in economic indicators since 2008-2009 (the great recession). To simplify, the variance

[12] with the exception of the M3 monetary aggregate and the index of sensitive materials price since they are only available up to 2004Q2.

parameters take back their value before 1987Q4. The model is given by:

$$y_t = \beta_0(S_t) + \sum_{i=1}^{r} \beta_{1,i}(S_t) B(L^{1/m}, \theta_i) \hat{f}_{i,t}^{(m)} + \delta \varepsilon_t(S_t) \qquad t = 1, \ldots, T \tag{16}$$

where δ is a positive parameter inferior to one over 1984Q2-2007Q4 and equal to one otherwise. We consider $r = 1$ factor driven by $q = 1$ dynamic shock and we assume that the factor follows an AR(1) process. We retain $K = 12$ so that the specification includes the last year of monthly data.

The estimation results of this model are given in Table 4. The link between the financial factor and the GDP dynamics is significant only in recessionary periods. The difference of the intercepts across the regimes is large. We also find the classical features of the business cycle: the expansion state is longer-lived and more volatile. As broadly reported in the previous literature, the break parameter is significant. The great moderation is characterized by an overall volatility divided per 5.

Figure 5 depicts the estimated weight function. We find a hump shaped function. According to this chart, the financial factor contains useful information to predict the US business cycle up to 6 months in advance. The weighting function peaks at the third month meaning that the financial factor contains particularly relevant information on the business cycle at this horizon. Note also that the function is very sharp. This is a favorable configuration to the MIDAS specification according to the results of the Monte Carlo simulations reported in the previous section.

The adjustment quality of the six competing models (ML1-ML3) and (MS1-MS3) is compared over the period 1959-2010 in Table 5. The results show that the MS-FaMIDAS model fits the US data more accurately and identifies better the states of the economy. Incorporating regime switching coefficients improves the fit of the models (MS versus ML models). Sizeable gain is also achieved when using higher frequency data, even in the linear framework (ML3 versus ML1 and MS3 versus MS1). The QPS criterion is better in MS3 and the turning points are much better identified when incorporating mixed frequency according to the TPI criteria. Finally, the parsimonious MS3 specification also performs better than the MS-MIDAS model with unrestricted lag polynomials MS2 in terms of regime estimation.

Figure 6 shows the smoothed probabilities of being in recession according to the MS-FaMIDAS model. The grey areas represent NBER recession periods. The model detects successfully the eight recessions over the period 1959-2010, even though the signals given for the two first ones remain weak. Note also two false signals: in 2002Q4 and in the

Table 4: Results of estimation of the MS-FaMIDAS on US GDP

p_{11}	p_{22}	$\beta_{0,1}$	$\beta_{1,1}$	$\beta_{0,2}$	$\beta_{1,2}$	σ_1	σ_2	δ
0.67	0.94	-0.03	-0.32	0.89	0.02	0.46	1.08	0.16
[2.13]	[3.99]	[-0.25]	[-3.12]	[17.88]	[0.53]	[2.52]	[6.11]	[4.49]

This table reports the parameter estimations and the associated t-statistics in brackets.

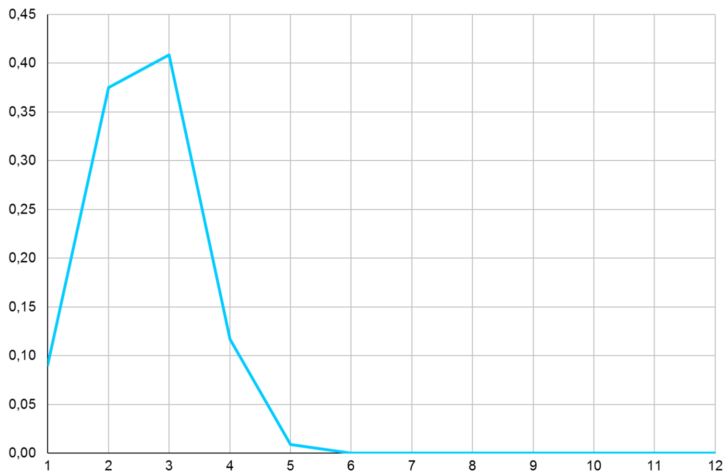


Figure 5: Estimated weights of the MS-FaMIDAS model

Table 5: In-sample performance of the six specifications on simulated data on US GDP

	k	R2	BIC	QPS	TPI (0.5)	TPI (0.4)
AR(1)	2	0.11	526.6	-	-	-
ML1	5	0.05	559.3	-	-	-
ML2	13	0.08	595.6	-	-	-
ML3	4	0.07	550.3	-	-	-
MS1	14	0.24	530.6	0.103	25%	31%
MS2	30	0.31	625.5	0.140	25%	31%
MS3	10	**0.31**	**517.8**	**0.079**	**68%**	**75%**

This table reports the R-squared, BIC, QPS and TIP criteria for the six specifications. k denotes the number of parameters estimated in each model. The TIP is computed alternatively with $\lambda = 0.5$ and $\lambda = 0.4$ and with a lead/lag parameter τ equal to two quarters.

second semester of 2006. The first one is probably related to the sharp drop in stock prices at the end of 2002 in stock exchanges across the United-States, Canada, Asia and Europe. Nevertheless, as shown by the TPI criterion, the MS-FaMIDAS detects more accurately the 16 turning points in the sample.

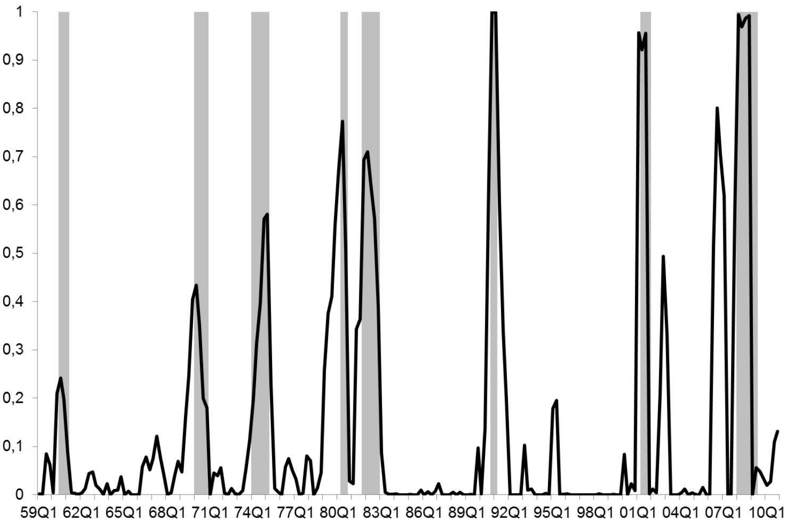


Figure 6: The probabilities of being in recession according to the MS-FaMIDAS

3.3 Out-of-sample results

We finally assess the quality of the forecasts made 7 months ahead to 1 month before the GDP release.

The evaluation is conducted in real conditions. First, the models are estimated from the observations available at the time of the forecast, using the vintages of output growth provided by Croushore and Stark (the financial variables are supposed to be not subject to data revisions). Moreover, the parameters of the models are estimated recursively using the only information available at the time of the forecast. The models are estimated from 1959 and the out-of-sample period spans from 1990Q1 to 2010Q4 which includes three recessions. Rather than using a balanced dataset, we also replicate the pattern of missing values at the end of the sample to take into account the time of publication of the variables, given in Appendix. The interpolation of the missing values at the end of the sample is carried out with the Kalman filter as explained in section 1.

More precisely, the approach is as follows. The first quarter of 1990 is forecast conditional on the information available in September 1989, October 1989, and so on up to March 1990 (the US quarterly GDP is release about one month after the end of the

Table 6: Out-of-sample performance of the six specifications on US data

		Direct approach							Iterative approach						
h		2	5/3	4/3	1	2/3	1/3	0	2	5/3	4/3	1	2/3	1/3	0
	AR	1.00	0.99	0.99	0.99	0.96	0.96	0.96	1.00	0.99	0.99	0.99	0.96	0.96	0.96
R	ML1	1.11	1.11	1.09	1.09	1.11	1.15	1.04	1.10	1.11	1.10	1.09	1.11	1.11	1.04
M	ML2	1.14	1.15	1.19	1.11	1.19	1.24	1.11	1.09	1.11	1.13	1.09	1.16	1.17	1.11
S	ML3	1.09	1.13	1.09	1.07	1.05	1.09	0.97	1.00	1.05	1.02	1.05	1.02	0.97	0.97
E	MS1	0.99	0.95	0.99	0.98	0.96	1.04	0.90	0.98	0.96	0.96	0.95	0.94	0.98	0.90
	MS2	1.00	0.97	1.04	0.99	0.94	1.04	0.90	0.98	0.95	0.98	0.96	0.97	0.97	0.90
	MS3	1.02	1.00	0.98	1.03	0.92	0.94	0.93	0.99	0.98	1.01	0.96	0.97	0.88	0.97
Q	MS1	0.22	0.16	0.16	0.15	0.11	0.11	0.10	0.14	0.14	0.13	0.13	0.11	0.11	0.10
P	MS2	0.17	0.18	0.17	0.17	0.13	0.14	0.12	0.14	0.17	0.16	0.14	0.15	0.13	0.12
S	MS3	0.14	0.15	0.15	0.13	0.12	0.11	0.09	0.14	0.12	0.13	0.12	0.09	0.10	0.09
TPI	MS1	0.50	0.67	0.67	0.67	0.83	0.83	0.83	0.17	0.67	0.67	0.67	0.83	0.83	0.83
(0.5)	MS2	0.17	0.67	0.33	0.67	0.83	0.83	0.83	0.50	0.67	0.67	0.50	0.50	0.57	0.83
	MS3	0.17	0.67	0.50	0.57	1.00	0.83	0.83	0.33	0.50	0.50	0.50	0.83	0.83	0.83
TPI	MS1	0.50	0.50	0.67	0.67	0.83	0.83	0.83	0.33	0.50	0.50	0.50	0.83	0.83	0.83
(0.4)	MS2	0.83	0.50	0.50	0.67	0.83	0.83	0.83	0.33	0.50	0.67	0.50	0.67	0.50	0.83
	MS3	0.50	0.83	0.33	0.83	1.00	0.83	0.83	0.33	0.67	0.67	0.50	0.83	0.83	1.00

This table reports the ratios of the RMSFEs relative to the RW model, the QPS and TPI criteria. The TPI criterion is provided for $\lambda = \{0.4, 0.5\}$ and $\tau = 2$.

reference quarter, that is in the specific case of the first quarter of 1990 in April 1990). The MS-FaMIDAS model is then estimated on these data-sets and used for prediction. Similarly, we produce seven forecasts of the GDP growth rate in the second quarter of 1990 from the available data in December 1990 to June 1990. These calculations are replicated up to the last quarter of the out-of-sample period.

We finally get 7 sets of forecasts of the GDP growth rate and of the occurrence of the recession state for the quarters 1990Q1 to 2010Q4. The GDP forecasts are compared to the final value of GDP approximated by the series available in September 2011 and the chain forecast is compared to the NBER datation. As done previously, the forecast accuracy of the GDP growth is measured with the RMSFE and the ability of the model to detect the turning points is assessed with the QPS and TPI indices.

Table 6 reports the ratios of RMSFE of the six specifications against the RW benchmark (a ratio below one indicates a gain relative to the naive forecast). To assess the quality of the forecast of the chain, we also provide the QPS and the TPI criteria for the three MS models. We compare the six specifications and the two usual benchmarks: AR and RW models. We also distinguish the results obtained with the direct and iterative approaches.

As far as the forecast of GDP growth is concerned, the nonlinear models perform better than the linear ones. However, we do not always find a gain when taking into account the mixed frequencies with MIDAS specifications. There is a gain relative to the random walk showing that the financial factor provides useful information to forecast US GDP growth. However, the quality of forecasts of the GDP growth rate is relatively poor since the six specifications do not beat the AR(1) model in many cases. Regarding the detection of recessions, the results are more satisfactory. In the direct and iterative approaches, the smallest QPS criteria are obtained with the constrained MS-FaMIDAS model. The MS-FaMIDAS specifications also detect a higher proportion of the observed turning points according to the TPI criterion. All specifications provides better signals when the forecast horizon shortens.

4 Concluding remarks

In this paper, we have introduced the MS-factor MIDAS model. This specification deals with several issues specific to short run forecasting. It allows exploiting the information provided by a large data-set and deals with mixed frequency variables. In addition, it gives quantitative and qualitative information about the state of the economy.

Monte Carlo evidence suggests that our estimation procedure provides robust estimates of the parameters of the model. The Monte Carlo experiments also show that the MS-FaMIDAS represents a robust forecasting device for various settings. We find a significant loss, both in-sample and out-of-sample, when ignoring the regime switches and the mixed frequencies if present in the data. In line with previous papers about MIDAS model, we also show that the MIDAS specification with unrestricted lag polynomials performs similarly than MIDAS model with constrained lags at least when the difference in sampling frequencies is small.

In the empirical application, we find that the MS-FaMIDAS model provides a better fit of the US GDP growth rate than linear specifications when using the information provided by monthly financial indicators over 1959-2010. It also detects more successfully the 8 recessions than specifications which time aggregates the high frequency indicators. The findings are less clear-cut when looking at out-of-sample results. The quality of forecasts of the GDP growth rate is relatively poor. Nevertheless, the MS-FaMIDAS model gives more accurate forecast of the recession state.

There are a number of potential extensions of this paper. In particular, it would be interesting to allow a switch in the parameters of the weighting function and time-varying probabilities. This is left for future research.

References

Andreou, E., E. Ghysels, and A. Kourtellos, Forecasting with mixed-frequency data, in M.P. Clements and D.F. Hendry, editors, *Oxford Handbook on Economic Forecasting*, 2010.

Armesto, M.T., K.M. Engemann, and M.T. Owyang, *Forecasting with mixed frequencies*, Federal Reserve Bank of Saint Louis Review 2010.

——, R. Hernández-Murillo, M.T. Owyang, and J. Piger, Measuring the information content of the Beige Book: a mixed data sampling approach, *Journal of Money, Credit and Banking*, 2010, *41* (1), 35–55.

Bai, J. and S. Ng, Forecasting economic time series using targeted predictors, *Journal of Econometrics*, 2008, *146* (2), 304–317.

——, E. Ghysels, and J. Wright, *State space models and MIDAS regressions*, Working Paper, NY Fed, UNC and John Hopkins 2009.

Banbura, M. and G. Runstler, A look into the factor model black box: Publication lags and the role of hard and soft data in forecasting GDP, *International Journal of Forecasting*, 2011, *27* (2), 333–346.

Boivin, J. and S. Ng, Are more data always better for factor analysis, *Journal of Econometrics*, 2006, *132* (1), 169–194.

Chauvet, M., An econometric characterization of business cycle dynamics with factor structure and regime switching, *International Economic Review*, 1998, *39* (4), 969–996.

Chevillon, G. and D.F. Hendry, Non-parametric direct multi-step estimation for forecasting economic processes, *International Journal of Forecasting*, 2005, *21* (2), 201–218.

Clements, M.P., A.B. Galvao, and J.H. Kim, Quantile forecasts of daily exchange rate returns from forecasts of realized volatility, *Journal of Empirical Finance*, 2008, *15* (4), 729–750.

—— and ——, Macroeconomic forecasting with mixed-frequency data, *Journal of Business and Economic Statistics*, 2008, *26* (4), 546–554.

—— and ——, Forecasting US output growth using leading indicators: an appraisal using MIDAS models, *Journal of Applied Econometrics*, 2009, *24* (7), 1187–1206.

Croushore, D., Frontiers of Real-Time Data Analysis, *Journal of Economic Literature*, 2011, *49* (1), 72–100.

—— and T. Stark, A real-time data set for macroeconomists, *Journal of Econometrics*, 2001, *105*, 111–130.

Diebold, F.X. and G.D. Rudebusch, Measuring business cycles: A modern perspective, *Review of Economics and Statistics*, 1996, *78* (1), 67–77.

Doz, C., D. Giannone, and L. Reichlin, A two-step estimator for large approximate dynamic factor models based on Kalman filtering, *Journal of Econometrics*, 2011, *164* (1), 188–205.

Foroni, C., M. Marcellino, and C. Schumacher, *U-MIDAS: MIDAS regressions with unrestricted lag polynomials*, Deutsche Bundesbank working paper 35 2011.

Ghysels, E., A. Sinko, and R. Valkanov, MIDAS regression: Further results and new directions, *Econometric Reviews*, 2007, *26* (1), 53–90.

—— , P. Santa-Clara, and R. Valkanov, *The MIDAS touch: mixed data sampling regression models*, Cirano working paper 2004.

—— , —— , and —— , There is a risk-return trade-off after all, *Journal of Financial Economics*, 2005, *76*, 509–548.

—— , —— , and —— , Predicting volatility: getting the most out of return data sampled at different frequencies, *Journal of Econometrics*, 2006, *131*, 59–95.

Giannone, D., L. Reichlin, and D. Small, Nowcasting: the real-time informational content of macroeconomic data, *Journal of Monetary Economics*, 2008, *55* (4), 665–676.

Guérin, P. and M. Marcellino, *Markov-Switching MIDAS models*, EUI working paper 2011-03 2011.

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

Kim, C.J. and C. Nelson, 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*, 1998, *80* (2), 188–201.

——— and ——— , Has the U.S. Economy Become More Stable? A Bayesian Approach Based on a Markov-Switching Model of the Business Cycle, *Review of Economics and Statistics*, 1999, *81* (4), 608–616.

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

Koenig, E.F., S. Dolmas, and J. Piger, The use and abuse of real-time data in economic forecasting, *Review of Economics and Statistics*, 2003, *85*, 618–628.

Kuzin, V.M., M. Marcellino, and C. Schumacher, MIDAS versus mixed-frequency VAR: nowcasting GDP in the euro area, *International Journal of Forecasting*, 2011, *27* (2), 529–542.

Marcellino, M. and C. Schumacher, Factor-MIDAS for nowcasting and forecasting with ragged-edge data: a model comparison for German GDP, *Oxford Bulletin of Economics and Statistics*, 2010, *72* (4), 518–550.

——— , J.H. Stock, and M.W. Watson, A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time-series, *Journal of Econometrics*, 2006, *135* (1-2), 499–526.

McConnell, M. M. and G. Perez-Quiros, Output Fluctuations in the United States: What Has Changed Since the Early 1980's?, *The American Economic Review*, 2000, *90* (5), 1464–1476.

Stock, J.H. and M.W. Watson, *Implications of Dynamic Factor Models for VAR Analysis*, NBER Working Paper 11467 2005.

Appendix: Data sources and descriptions

code	variable	transf	delay	datastream code
FM1	M1	6	1	USM1....B
FM2	M2	6	1	USM2....B
FM2DQ	real M2	5	1	USM2....D
FMFBA	monetary base, adjusted for reserve requirement changes	6	0	USM0....B
FMRRA	depository inst reserves:total,adj for reserve req chgs	6	0	USTOTRSAB
FMRNBA	depository inst reserves:nonborrowed, adj for reserve req chgs	6	0	USNBRRSAB
FCLNQ	commercial - industrial loans oustanting in 1996 dollars	6	1	USFCILO.C
FCLBMC	weekly RP LG com'l banks: net change com'l and industrial loans	1	1	USCILNNCB
CCINRV	consumer credit outstanding - nonrevolving	6	1	USCRDCONB
A0M095	Ratio, consumer installment credit to personal income	2	1	USCSCRE%Q
FSPCOM	Standard and Poor's index	5	0	USS PCOM
DJPIN	US Dow Jones Industrials share prince index	5	0	USSHRPRCF
FSDXP	Standard and Poor's: dividend yield	2	0	USSPDIVY
FSPXE	Standard and Poor's: price-earnings ratio	5	0	USSPRPER
FYFF	interest rate: federal funds (effective)	2	0	USFDFUND
CP90	Commercial Paper Rate (AC)	2	0	USI60BC.
FYGM3	interest rate: U.S.treasury bills,SEC MKT,3-MO	2	0	USGBILL3
FYGM6	interest rate: U.S.treasury bills,SEC MKT,6-MO	2	0	USYTB6SM
FYGT1	interest rate: U.S.treasury const maturities,1-YR	2	0	USTRCN1.
FYGT5	interest rate: U.S.treasury const maturities,5-YR	2	0	USTRCN5.
FYGT10	interest rate: U.S.treasury const maturities,10-YR	2	0	USTRCN10
FYAAAC	bond yield: Moody's AAA corporate	2	0	USCRBYLD
FYBAAC	bond yield: Moody's BAA corporate	2	0	USCRBBAA
scp90	cp90-fyff	1	0	ok
sfygm3	fygm3-fyff	1	0	ok
sFYGM6	fygm6-fyff	1	0	ok
sFYGT1	fygt1-fyff	1	0	ok
sFYGT5	fygt5-fyff	1	0	ok
sFYGT10	fygt10-fyff	1	0	ok
sFYAAAC	fyaaac-fyff	1	0	ok
sFYBAAC	fybaac-fyff	1	0	ok
EXRUS	US effective exchange rate	5	0	USI..NEUE
EXRSW	exchange rate (Swiss Franc/USD)	5	0	SWXRUSD.
EXRJAN	exchange rate (Yen/USD)	5	0	JPXRUSD.
EXRUK	exchange rate (UK/USD)	5	0	UKXRUSD.
EXRCAN	exchange rate (CAN/USD)	5	0	CNXRUSD.
PWFSA	PPI: finished goods	6	1	USPROPRCE
PWFCSA	PPI: finished consumer goods	6	1	USWPCONFE
PWIMSA	PPI: intermed mat.supplies and components	6	1	USWPINTME
PWCMSA	PPI: crude materials	6	1	USWPCRUDE
PXOIL	average brent oil price	6	1	UKOILBREN
PMCP	NAPM commodity price index	1	1	ISM
PUNEW	CPI: all items	6	1	USCONPRCE
PU83	CPI: apparel and upkeep	6	1	USCPAPPLE
PU84	CPI: transportation	6	1	USCPTRANE
PU85	CPI: medical care	6	1	USCPMEDCE
PUC	CPI-U: commodities	6	1	USCPCOMME
PUCD	CPI-U: durables	6	1	USCPD...E
PUS	CPI-U: services	6	1	USCPSERVE
PUXF	CPI-U: all items less food	6	1	USCPXF..E
PUXHS	CPI-U: all items less shelter	6	1	USCPXHS.E
PUXM	CPI-U: all items less medical care	6	1	USCPXMEDE
GMDC	PCE, implicit price deflator	6	1	USUUBA8BE
GMDCD	PCE, implicit price deflator; durables	6	1	USUND7JGE
GMDCN	PCE, implicit price deflator; non durables	6	1	USUBN11JE
GMDCS	PCE, implicit price deflator; services	6	1	USU4H9R7E

Note: The column transf gives the transformation of the variables taken in level (code 1), first difference (2), second difference (3), log-level (4), log-first-difference (5) or log-second-difference (6).

Documents de Travail

370. G. Verdugo, H. Fraisse et G. Horny, “ Évolution des Inégalités Salariales en France : le Rôle des Effets de Composition,” Mars 2012
371. N. Berardi, Th. A. Eife and E. Gautier, “ Optimal Price Setting During a Currency Changeover: Theory and Evidence from French Restaurants,” March 2012
372. Ch. A. E. Goodhart, A. K Kashyap, D. P. Tsomocos and A. P. Vardoulakis, “Financial Regulation in General Equilibrium,” March 2012
373. M. Bussière, M. Hoerova and B. Klaus, “Commonality in Hedge Fund Returns: Driving Factors and Implications,” March 2012
374. C. Glocker and P. Towbin, “The macroeconomic effects of reserve requirements,” April 2012
375. E. Gautier and R. Le Saout, “The dynamics of gasoline prices: evidence from daily french micro data,” April 2012
376. A. Banerjee and S. Malik, “The changing role of expectations in US monetary policy: A new look using the Livingston Survey,” April 2012
377. H. Pagès, “Bank monitoring incentives and optimal ABS,” April 2012
378. H. Pagès and D. Possamai, “A mathematical treatment of bank monitoring incentives,” April 2012
379. P. Fève, J. Matheron and J.G. Sahuc, “A pitfall with DSGE-Based, estimated, government spending multipliers,” April 2012
380. M. Boutillier and J. C. Bricongne, “Disintermediation or financial diversification? The case of developed countries,” April 2012
381. Y. Ivanenko and B. Munier, “Price as a choice under nonstochastic randomness in finance,” May 2012
382. L. Agnello and R. M. Sousa, “How does Fiscal Consolidation Impact on Income Inequality?,” May 2012
383. L. Ferrara, M. Marcellino, M. Mogliani, “Macroeconomic forecasting during the Great Recession: The return of non-linearity?,” May 2012
384. M. Bessec and O. Bouabdallah, “Forecasting GDP over the business cycle in a multi-frequency and data-rich environment,” June 2012

Pour accéder à la liste complète des Documents de Travail publiés par la Banque de France veuillez consulter le site : www.banque-france.fr

For a complete list of Working Papers published by the Banque de France, please visit the website: www.banque-france.fr

Pour tous commentaires ou demandes sur les Documents de Travail, contacter la bibliothèque de la Direction Générale des Études et des Relations Internationales à l'adresse suivante :

For any comment or enquiries on the Working Papers, contact the library of the Directorate General Economics and International Relations at the following address :

BANQUE DE FRANCE
49- 1404 Labolog
75049 Paris Cedex 01
tél : 0033 (0)1 42 97 77 24 ou 01 42 92 63 40 ou 48 90 ou 69 81
email : 1404-ut@banque-france.fr