DOCUMENT

DE TRAVAIL

N° 520

VARIABLE SELECTION IN PREDICTIVE MIDAS MODELS

Clément Marsilli

November 2014

BANQUE DE FRANCE

EUROSYSTÈME

DIRECTION GÉNÉRALE DES ÉTUDES ET DES RELATIONS INTERNATIONALES

DIRECTION GÉNÉRALE DES ÉTUDES ET DES RELATIONS INTERNATIONALES

VARIABLE SELECTION IN PREDICTIVE MIDAS MODELS

Clément Marsilli

November 2014

Variable Selection
in Predictive MIDAS Models

Clément Marsilli*

I would like to thank Juan-Pablo Ortega, Laurent Ferrara, Gong Cheng, Arnaud Dufays, Eric Ghysels, Domenico Giannone, Claude Lopez, Daniele Siena, Julia Schmidt, Luc Bauwens, Frederic Bec, Massimiliano Marcellino, Christian Schumacher, as well as seminar participants at the EEA-ESEM 2013 meeting, at the SAEe 2013 meeting, at the ADRES 2014 meeting, at the Deutsche Bundesbank and at the Banque de France for very helpful comments. The views expressed herein are solely those of the author and do not reflect the views of the Banque de France or the Eurosystem.

*Banque de France, International Macroeconomics Division, clement.marsilli@banque-france.fr.

Résumé

Dans le cadre de prévisions de court terme de l'activité économique, mobiliser l'ensemble de l'information disponible est un élément essentiel à une bonne modélisation. La différence dans les fréquences d'échantillonnage des indicateurs macroéconomiques restreint un usage efficient de ces données. A cet égard, la modélisation multi-fréquentielle MIDAS (*Mixed-Data Sampling*) a prouvé sa capacité à agréger de manière optimale les séries temporelles pour la prévision économique. Le choix des variables à privilégier demeure néanmoins délicat. Cet article entend développer différentes techniques de sélection de variables à fréquences multiples dans l'optique d'améliorer la capacité prédictive de la modélisation MIDAS. Deux méthodes basées respectivement sur la sélection par régression pénalisée et la recherche stochastique bayésienne sont notamment introduites dans cette étude. Celles-ci sont combinées à une procédure de validation croisée permettant une sélection automatique des variables selon leur apport prédictif. Enfin, les aptitudes empiriques des différentes stratégies MIDAS sont comparées *via* un exercice de prévision du PIB américain sur la période 2000-2013 à partir de données journalières et mensuelles. Il apparaît que les modélisations développées dans cet article parviennent à identifier les indicateurs conjoncturels avancés et améliorent ainsi leur qualité prédictive.

Mots-clés – Prévisions du PIB, Modélisation multi-fréquentielle, MIDAS, Sélection de variables.

Abstract

In short-term forecasting, it is essential to take into account all available information on the current state of the economic activity. Yet, the fact that various time series are sampled at different frequencies prevents an efficient use of available data. In this respect, the *Mixed-Data Sampling* (MIDAS) model has proved to outperform existing tools by combining data series of different frequencies. However, major issues remain regarding the choice of explanatory variables. The paper first addresses this point by developing MIDAS based dimension reduction techniques and by introducing two novel approaches based on either a method of penalized variable selection or Bayesian stochastic search variable selection. These features integrate a cross-validation procedure that allows automatic in-sample selection based on recent forecasting performances. Then the developed techniques are assessed with regards to their forecasting power of US economic growth during the period 2000-2013 using jointly daily and monthly data. Our model succeeds in identifying leading indicators and constructing an objective variable selection with broad applicability.

Keywords – Forecasting, Mixed frequency data, MIDAS, Variable selection, GDP

JEL Codes – C53, E37.

Non-technical summary

Recent economic events have cast serious doubts on standard methods of economic prediction which failed to provide accurate economic snapshot and forecast for policymaking in time of crisis. In fact, econometric forecasting models are usually based on regressions that seek to explain and predict economic growth through a range of contemporaneous and historical information. However, the volume of available data for economic forecasting is huge. An important achievement would be to determine the more relevant set of information coming from industrial figures, employment statistics, opinion surveys, prices of commodities, stocks, bonds quoted in *quasi*-continuous time, indicators of real estate market, etc. These time series can potentially be explanatory variables for economic prediction but they also can contain noise. To only extract relevant information from these variables is as difficult as to separate the wheat from the chaff. Time series coming from real and financial economy do not have the same characteristics, both in terms of sampling frequency and predictive power. The financial data are especially generally less frequently used in macroeconomic forecasting models. Indeed, financial market data are inherently volatile and therefore it is difficult to discern their real predictive input from a macroeconomic point of view. In addition, the temporal frequency of financial data is almost continuous and hence difficult to use in the context of a quarterly modeling.

The Mixed-Data Sampling (MIDAS) methodology has been developed to suit these purposes and moreover has proved to outperform existing tools by combining data series of different frequencies. In fact, the underlying idea is simple but innovative; weight coefficients that allow the temporal aggregation of high-frequency data rely on a function with a small number of parameters. Those are estimated as part of the whole optimization process. Such data-driven methodology exploits the predictive ability of our set of information to improve empirical results. For macroeconomic forecasting, how to choose explanatory variables is tricky and heavily depends on the forecast horizon. There is a dilemma regarding this choice: on the one hand we want to include as many explanatory variables as possible to increase the predictive power of the model; on the other hand, facing the limited sample size, we need to care about model sparsity. This paper focuses on variable selection techniques in mixed-frequency models using both daily and monthly data for short-term forecasting. More specifically, we address the selection issue by developing MIDAS-based dimension reduction techniques and introducing two new methods using either a method of penalized variable selection, (i) the LASSO augmented MIDAS model, or (ii) the Bayesian stochastic search variable selection. These novel strategies are then compared with (iii) the Factor-Augmented MIDAS (FAMIDAS) model, and (iv) a forecasts combination technique of univariate MIDAS based predictions. In these four approaches, the selection is automatically carried out in-sample using a cross-validation procedure based on recent forecasting performances. We empirically assess the different selection methods by comparing point forecasts on the US GDP growth from 2000 to 2013. We show that the set of chosen predictors determined by the proposed variable selection procedure reflects the varying nature of the economic outlook. An adequate variable selection significantly improves forecasting performances for all phases of the business cycle observed.

1 Introduction

Short-term analysis aims at providing forecasts based on all the available information, it is usually required to use data sampled at different frequencies. The Great Recession experienced by the main industrialized countries during the period 2008-2009 in the wake of the American subprimes crisis has encouraged many forecasters to reconsider their model specifications, especially regarding the interactions between financial and macroeconomic variables. In this respect, Foroni and Marcellino (2014) and Banbura et al. (2012) have recently reviewed the existing mixed-frequency models designed for handling immediate past data (usually referred to as *ragged edge* data) and for nowcasting.

In this paper, we focus on the Mixed Data Sampling (MIDAS) method that allows the use of high frequency series to explain low frequency variables. Introduced by Ghysels et al. (2002) and more formally conceptualized by Ghysels et al. (2007) and by Andreou et al. (2010), the MIDAS approach constitutes a parsimonious weighting framework for handling distributed lags. This method allows us to explain a low frequency variable by using exogenous variables sampled at higher frequencies without resorting to any aggregation procedure. It has proved to be particularly suitable in macroeconomic forecasting and in capturing early signals of turning points using multifrequency explanatory variables (see for example Ferrara and Marsilli, 2013). Furthermore, the MIDAS regression has been used to predict quarterly GDP fluctuations using both monthly real economic data and daily financial series; Andreou et al. (2013) and Ferrara et al. (2014) showed that this combination of information significantly improves the prediction results. Many works prove that an appropriate selection of explanatory variables, regardless of their sampling frequencies, has a major impact on the performance of this forecasting method. In macroeconomic forecasting, empirical models are generally based on dimension reduction methods coming from either variable or model selection. The difference between these two close schemes is mainly methodological: variable selection aims at an a priori determination of the relevant predictors, while model selection provides an algorithmic approach to combine models which are typically univariate. *The main goal of this paper is introducing and comparing various variable selection methods within the mixed-frequency framework for macroeconomic forecasting. We will show that the use of well targeted predictors significantly improves the quality of forecasts. All schemes proposed have their grounds on model selection to tackle the so-called curse of dimensionality within a MIDAS forecasting framework.*

A large family of widely used techniques in the literature for economic forecasting is based on principal component analysis and factor models. We refer, among others, to Forni et al. (2000) or Stock and Watson (2002). In the context of mixed-frequency models, Marcellino and Schumacher (2010) have put forward a dynamic factor MIDAS model (FAMIDAS) as a way to tackle the lack of parsimony associated to the profusion of covariates. FAMIDAS is a method to incorporate in a MIDAS framework standard tools of factor analysis that usually produce very good results for short-term forecasting (see Giannone et al., 2008 or Barhoumi et al., 2010) and hence represents a competitive benchmark when we compare the performance of different models. As an alternative to principal components analysis De Mol et al. (2008) have suggested the use of Bayesian regressions or penalized regressions (especially LASSO method[1] introduced by Tibshirani, 1996) as a dimension reduction technique.

Other approaches for using mixed frequency data are the bridge models introduced, for instance, by Barhoumi et al. (2008) for forecasting purposes. In Bencivelli et al. (2012) bridge based techniques are put together with Bayesian Model Averaging (BMA) to combine predictions coming from various model settings. The literature shows that forecast combinations and, in particular, model averaging like BMA, yield good forecasting results. Indeed, Palm and Zellner (1992) claim that "*in some instances it is sensible to use a simple average of individual forecasts*". Rodriguez and Puggioni (2010) have recently adapted Bayesian approaches to estimate MIDAS models for forecasting exercises. In their paper, BMA provides a way to estimate the weights applied on the explanatory variables. Unfortunately, searching the best model using this approach involves maximizing marginal likelihoods and hence requires in general the assessment of the 2^n different combinations of models which may prove to be numerically expensive. Another technique available in the literature that extends the Bayesian selection analysis to stochastic search relies on the mixture of priors on regressor coefficients with spike and slab components. In this respect, we refer to Mitchell and Beauchamp (1988) for Dirac point mass spikes or to George and McCulloch (1993) for absolutely continuous spikes. These techniques have been widely exploited in econometrics; see, for example, Korobilis (2013) for an empirical application to the prediction of economic growth or Kaufmann and Schumacher (2012) for finding sparsity on factor models. The recent paper by Scott and Varian (2013) also considers spike and slab regression for variable selection in the nowcasting of economic time series.

In this paper, we will focus on four different dimension reduction techniques that we combine with the MIDAS regression structure. More specifically, we introduce two new methods: (i) the LASSO augmented MIDAS model and (ii) the Bayesian MIDAS model with stochastic search variable selection. These novel strategies are then compared with (iii) the Factor-Augmented MIDAS (FAMIDAS) model, and (iv) a forecasts combination technique of univariate MIDAS based predictions. In these four approaches, the selection is carried out in-sample using a cross-validation procedure based on recent forecasting performances. We empirically assess the different selection methods by comparing point forecasts and prediction errors on the US GDP growth from 2000 to 2013. Our empirical results allow us to draw several important conclusions: first, we show that adequate variable selection significantly improves forecasting performances for all phases of the business cycle observed. Second, we observe that the two novel techniques developed succeeded in identifying early signals of the Great Recession from 3 to 6 months in advance while two other models were unable to capture the downturn. Third, the set of chosen predictors determined by the proposed variable/model selection procedure reflects the varying nature of the economic outlook.

The paper is structured as follows: Section 2 presents first the general MIDAS regression framework. Section 3 and Section 4 exhibits novel variable selection techniques that we develop in this paper, respectively the LASSO augmented MIDAS model and the Bayesian MIDAS with

[1]LASSO, namely Least Absolute Shrinkage and Selection Operator, is described in detail in Section 3.

stochastic search technique. In Section 5, we introduce the predictive cross-validation selection strategy. Then, in Section 6, we empirically show how the proposed selection methods of explanatory variables out of a universe of well-known economic variables can significantly improve short-term forecasts of US GDP.

2 MIDAS regression framework

A standard way to proceed when dealing simultaneously with daily, monthly, and quarterly macroeconomic variables is to temporally aggregate higher frequency variables in order to homogenize their sampling frequency with respect to the lowest one. Since this approach carries intrinsically in its wake a loss of information, we prefer using the MIDAS technique put forward by Ghysels et al. (2002) in order to directly accommodate variables with different sampling frequencies without aggregating the available information. This section starts with a brief presentation of the MIDAS framework. Additionally, we present in detail the two variable selection techniques announced in the introduction using a LASSO type penalization scheme and Bayesian stochastic search.

We denote by y_t the explained variable of interest that we assume is sampled at the lowest possible frequency and let x_t be a vector of n time series whose components $x_{t,i}^{\kappa_i}$ are sampled with periods κ_i that divide that of y_t, that is $x_{t,i}^{\kappa_i}$ is quoted $1/\kappa_i$ times for each quote of y_t. The multivariate MIDAS regression model is specified by:

$$y_t = \beta_0 + \sum_{i=1}^{n} \beta_i \; m^{K_i}(\theta_i, L)\, x_{t,i}^{\kappa_i} \; + \varepsilon_t. \tag{1}$$

We emphasize that in equation (1) each explanatory variable can be sampled at a different frequency. This extension of the MIDAS regression has been proposed by Ghysels et al. (2007) and, recently, Andreou et al. (2013) and Ferrara et al. (2014) used that specification to combine daily financial variables with monthly indicators. The temporal aggregation takes place via the MIDAS regression kernel $m^K(\cdot)$ which smooths out the K past values of the variable $x_{t,i}$ by using a functional polynomial of the form

$$m^K(\theta, L) := \sum_{k=1}^{K} \frac{f(k,\theta)}{\sum_{l=1}^{K} f(l,\theta)} L^{(k-1)\kappa} \ , \tag{2}$$

where $L^{(k-1)\kappa}$ is the lag operator. For example, if we have $k=2$ and $\kappa = 1/3$, then $L^{(k-1)\kappa}x_t^{\kappa} = x_{t-1/3}$; this configuration corresponds to one lag of a monthly sampled indicator. As we could see in the previous expression, the MIDAS regression kernel is defined as a weighted average, where the weights are specified by a family of functions $f(\cdot)$ parametrized by θ. There exist different choices for this weight function available in the econometric literature involving either linear or nonlinear specifications. Ghysels et al. (2007) have presented a long list of possible forms that the

MIDAS function can take. In our study we choose to work with the two parameter exponential Almon Lag family:

$$f(k, \theta) := f(k, \theta_1, \theta_2) = \exp(\theta_1 k + \theta_2 k^2).$$

The parsimonious form of this MIDAS kernel allows us to limit the dimensionality of the regression problem at the time of aggregating high frequency. Clements and Galvão (2009) and Kuzin et al. (2011), among others, have shown that this kernel performs well in the prediction of quarterly economic growth with monthly indicators. We note that MIDAS is usually implemented in the literature with only a small number of variables because the nonlinearity of the weight function may cause difficulties in the estimation of the regression parameters. This fact shows the importance of developing variable selection methods, which is the subject of the following two sections.

3 The LASSO augmented MIDAS model

LASSO (Least Absolute Shrinkage and Selection Operator) has been introduced by Tibshirani (1996) as a covariates selection method in a linear regression setup. LASSO operates by penalizing the optimization problem associated to the regression with a term that involves the ℓ_1-norm of the coefficients. It belongs to the family of penalized regression model which amounts to performing least squares with some additional constraints on the coefficients, the ℓ_1-norm in the case of LASSO. Ng (2012) have shown that LASSO tends to have a lower misspecification risk in forecasting models when compared with usual information criteria. In the econometrics setup Bai and Ng (2008) and Schumacher (2010) have proposed to forecast economic series by using a combination of factor analysis with a LARS (see Efron et al., 2004) implementation of LASSO.

To be more specific, the LASSO takes advantage of the sparsifying properties of the ℓ_1-norm when solving the penalized optimization problem,

$$\begin{aligned} \widehat{b} &= \arg\min_b \sum_t \left(y_t - b_0 - \sum_i b_i \, x_{t,i} \right)^2 + \lambda_{\text{lasso}} \sum_i |b_i| \\ &= \arg\min_b ||Y - Xb||_2^2 + \lambda_{\text{lasso}} \, ||b||_1, \end{aligned} \tag{3}$$

where y_t is the dependent variable, x_t is the vector of covariates, b is the vector containing the regression parameters, and λ_{lasso} is the exogenous parameter which controls the strength of the LASSO penalization. The LASSO method does indeed reduce the dimension of the explanatory matrix X by driving non informative β_i elements to zero. Increasing $\lambda_{\text{lasso}} \in \mathbb{R}^+$ brings gradually elements of the β vector to zero, hence selecting relevant explanatory variables. The choice of the exogenous parameter λ_{lasso} that determines the number of covariates that are eliminated is essential and therefore a key issue that we will address later on via cross-validation.

Ridge regression is another popular penalized optimization scheme which, as opposed to the ℓ_1 penalty of LASSO, is based on a ℓ_2-norm penalty. Figure 1 illustrates the underlying principle of both techniques in the case of a multivariate regression model with two variables: b_1 and b_2. The LASSO is on the left, and the ridge regression on the right.

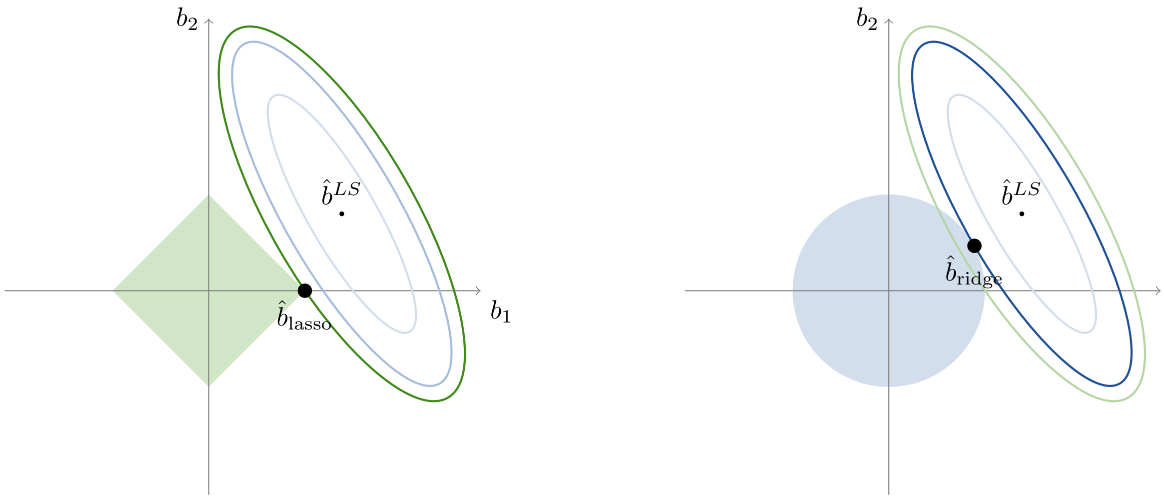

FIGURE 1: Penalized least squares estimate for the ℓ_1-norm (green) and the ℓ_2-norm (blue)

The ellipses around the least square estimator, $\hat{b}^{LS}$ represent the level sets of the squared error function $\|Y - Xb\|_2^2$ and the light colored areas correspond to balls of the ℓ_1 and ℓ_2 norms. In view of expression (3), the solution of the optimization problem that we are interested in takes place at the points in which both surfaces are tangent. The geometry of the problems makes that in the $\hat{b}^{\ell_1}$ case, the solution is generically located at the vertices of the ℓ_1-balls and hence the LASSO penalized solutions have entries equal to zero. The ridge based solutions are generally not located at that kind of specific points and are hence not necessarily sparse.

We put forward an extension of the LASSO model to the nonlinear MIDAS regression context by proposing the following optimization problem:

$$\begin{aligned}[\,\hat{\beta}, \hat{\theta}\,] &= \underset{\beta,\,\theta}{\arg\min} \sum_t \left(y_t - \beta_0 - \sum_{i=1}^{n} \beta_i \; m^{K_i}(\theta_i)\, x_{t,i}^{\kappa_i} \right)^2 + \lambda \sum_i |\beta_i| \\ &= \underset{\beta,\,\theta}{\arg\min} \, \|Y - \mathsf{X}(\theta)\,\beta\|_2^2 + \lambda\, \|\beta\|_1,\end{aligned} \tag{4}$$

where the matrix $\mathsf{X}(\theta)$ contains the MIDAS specifications that we previously described in (2),

$$\mathsf{X}(\theta) = \begin{pmatrix} 1 & m^{K_1}(\theta_1, L)\, x_{1,1}^{\kappa_1} & \cdots & m^{K_n}(\theta_n, L)\, x_{1,1}^{\kappa_n} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & m^{K_1}(\theta_1, L)\, x_{T,n}^{\kappa_1} & \cdots & m^{K_n}(\theta_n, L)\, x_{T,n}^{\kappa_n} \end{pmatrix}. \tag{5}$$

As we will see later on and like in the linear case, the ℓ_1 penalization on the β parameters implies a selection of the most relevant predictors. The number of covariates eliminated can be chosen by tuning the value of the exogenous parameter λ. A technical complication in solving (4) via any gradient descent method arises due to the non-smooth nature of the ℓ_1 norm. We overcome this difficulty using a local regularization technique due to Nesterov (2005). Further details are provided in the Appendix A.

As opposed to other standard iterative variable selection techniques, the combination of LASSO with the MIDAS regression presents the advantage to be a one-step procedure. This feature affects directly the numerical effort involved in its implementation, where the most expensive step is the determination of the penalization strength λ. This parameter will be selected using the *predictive cross-validation method.*

4 Bayesian variable selection in MIDAS models

Another approach that we explore in order to define the relevant subset of variables which should be included in the final regression model, we consider a specific Bayesian variable selection technique that relies on *spike and slab priors* (see George and McCulloch (1993)). This stochastic variable selection strategy is an a alternative to other usual Bayesian constructions that involve the comparison of all 2^n possible models , where n is the number of explanatory covariates under consideration. The approach that we propose yields a hierarchy on the covariates with respect to posterior distributions and relative inclusion probabilities. Kaufmann and Schumacher (2012) have recently used this technique to find relevant variables in sparse factor models.

The model selection relies on drawing the posterior ordinate using the Bayes formula. Indeed, we assume that residuals of the MIDAS regression model, as defined in (1), follow a Gaussian distribution $\mathcal{N}(0, \sigma)$. Thus, the conditional likelihood function of the MIDAS model under study has the following form:

$$f(Y|\beta, \theta, \sigma) = \frac{1}{(2\pi\sigma)^{T/2}} \exp\left[\frac{1}{2\sigma}(Y - \mathsf{X}(\theta)\beta)'(Y - \mathsf{X}(\theta)\beta)\right], \tag{6}$$

where $(Y - \mathsf{X}(\theta)\beta)$ stands for the matrix expression of the MIDAS regression (see equation (4)).

Bayesian approaches have been rarely used in the context of MIDAS regression model; the main reference in this direction is Rodriguez and Puggioni (2010) where the authors focus not on variable selection but on the number of temporal lags used in the regression. In that context, they use an exponential Almon weight function combined with linear methods that are reminiscent of the U-MIDAS scheme in Foroni et al. (2013).

In the Bayesian framework, model parameters are derived from the posterior density which is, according to the Bayes formula, proportional to the likelihood times the prior, that is,

$$\pi(\beta,\theta,\sigma|Y) = \frac{f(Y|\beta,\theta,\sigma)\times\pi(\beta,\theta,\sigma)}{f(Y)} \tag{7}$$

$$\propto f(Y|\beta,\theta,\sigma)\times\pi(\beta,\theta,\sigma). \tag{8}$$

In what follows, we exclusively focus on the equation (8).

Assuming that the parameters β, θ, and σ^2 are unknown, we define priors for each of these parameters. Considering the MIDAS model as a standard linear regression model that involves a nonlinear kernel which enables mixing frequencies, we use conjugate priors which provide a posterior distribution coming from the same family as the prior for the linear parameters. In fact, we suppose that priors for the regression parameters β are normally distributed and use Gamma prior for the MIDAS lag polynomial coefficients both θ_1 and θ_2 as suggested by Ghysels (2012). Furthermore, we choose specific priors that help us in determining whether a variable should be included or not. Indeed, we use with the spike and slab priors technique introduced by Mitchell and Beauchamp (1988) that constraints regressor coefficients to be zero (coefficient drawn from the "spike" prior) or not (drawn from the flat distribution: the "slab" prior). More specifically, we adopt a generalization of this method due to George and McCulloch (1993) and usually referred to as Stochastic Search Variable Selection (SSVS) that takes as prior the following mixture of two normal distributions:

$$\beta_i|h_i \sim h_i\mathcal{N}(0,\varphi^2) + (1-h_i)\mathcal{N}(0,c\,\varphi^2), \tag{9}$$

where c is a small positive number, φ^2 sufficiently large, and h_i is the binary random variable defined by

$$h_i = \begin{cases} 1 & \text{with} \quad \pi(h_i = 1) = \omega_i, \\ c & \text{with} \quad \pi(h_i = c) = 1-\omega_i, \end{cases} \tag{10}$$

which allows the switching from a density concentrated around zero to another one with larger variance. When $h_i = 1$, β_i exhibits flat distribution and therefore we can consider that the covariate $x_{t,i}^{\kappa_i}$ that goes with it should be included in the model; conversely, when $h_i = c$, the density of the coefficient is concentrated around the zero value. Figure 2 illustrates this mixture of both normal distributions.

We infer that, when $h_i = c$, corresponding variable should not be taken into account as a regressor. Consequently, formulas (9) and (10) can be interpreted by saying that ω_i is the prior probability that $x_{t,i}^{\kappa_i}$ should be kept as a explanatory variable. This particular feature of the SSVS method has been often reviewed in the literature. In particular, more complex choices of prior can be made: for example, George and McCulloch (1997) defined an hierarchical prior for the inclusion probability using a Beta distribution and Yuan and Lin (2005) have preferred the definition of a hierarchical Bayes formulation to show that it can be related to the LASSO estimator. Other references are Ishwaran and Rao (2005) or Malsiner-Walli and Wagner (2011).

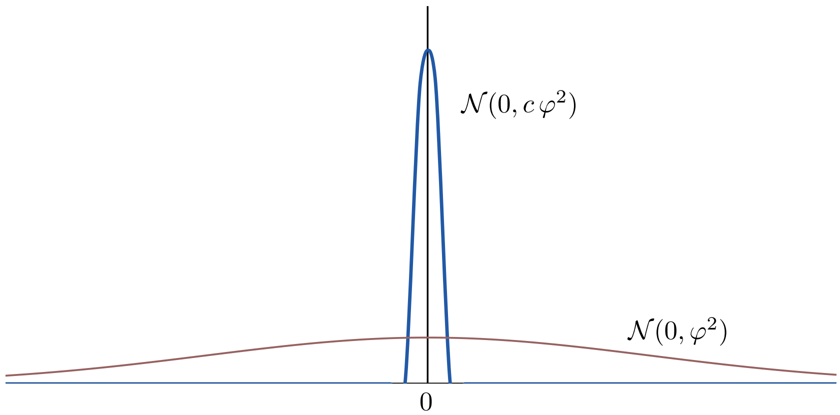

FIGURE 2: Mixture of a slab (red) and a spike (blue) normal distributions

In order to compute the posterior distribution, we proceed by implementing a Gibbs sampler to generate a ergodic Markov chain in which all parameters $(h, \omega, \beta, \theta, \sigma)$ are embedded. For the MIDAS parameter θ step, we use an Independence Chain Metropolis Hasting algorithm (iMH) within the Gibbs sampler to draw the posterior conditional distribution of θ. Details on the algorithm are provided in Appendix B. The algorithm converges relatively fast to a steady state of the Markov chain and the distribution obtained is an approximation of the posterior distribution which informs us about the selection that can be carried out in terms of the probabilities ω_i.

Finally, we establish a probability threshold $\Omega \in [0, 1]$ via a predictive cross-validation technique based on forecasting performance such that when $0 < \omega_i < \Omega < 1$ we consider that the relative predictor $x_{t,i}^{\kappa_i}$ should not be included in the model. We emphasize that in the same vein as the LASSO approach, the stochastic search variable selection yields a one-step estimation and selection procedure.

5 Predictive cross-validation

In this section, we propose a cross-validation method in order to determine model specifications that possess best predictive power. In this respect, we assume that the selection is updated according to its predictive error for the d previous. We investigate four families of forecasting models that we specify using the predictive cross-validation: the LASSO augmented MIDAS, the Bayesian-MIDAS Stochastic Search, the FAMIDAS, and the forecast combination of univariate MIDAS regressions.

We start by defining the variable ξ_i that be used as an indicator that determines whether the i^{th} variable must be taken into account or not in the model, that is,

$$\xi_i = \begin{cases} 1, & \text{if } x_{i,t}^{\kappa_i} \text{ is selected to be present in the model,} \\ 0, & \text{otherwise.} \end{cases}$$

We now rewrite the MIDAS model using the ξ_i variables and within the direct multistep forecasting frame, at the horizon h:

$$\hat{y}_{t+h|t} = \hat{\beta}_0 + \sum_{i=1}^{n} \xi_i \, \hat{\beta}_i \; m^{K_i}(\hat{\theta}_i, L) \, x_{t,i}^{\kappa_i}, \tag{11}$$

where $(\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_n, \hat{\theta}_1, \ldots, \hat{\theta}_n)$ are parameter estimates usually obtained by either non-linear least squares or maximum likelihood methods.

Notice that the effective size of the explanatory subset is determined by the ξ variables that take a non-zero value and that is chosen using the LASSO and the Bayesian SSVS based procedures adapted to the MIDAS context that we presented in the previous section. In order to evaluate their performance we take as a benchmark the Factor-Augmented MIDAS (FAMIDAS) model put forward by Marcellino and Schumacher (2010) and recently used by Ferrara and Marsilli (2014) to monthly nowcast the global economic growth. This approach is based on the combined use of two techniques: first, pooling information from blocks of covariates that share the same frequency into a certain number of factors, and second, tracking the dependent variable with a MIDAS regression model by incorporating these factors as explanatory variables.

Another benchmark that we also consider is forecast combination. Indeed, there is a growing volume of literature that shows that combining forecasts is a particularly competitive in prediction tasks; we refer to the very complete survey of Timmermann (2006). This technique has also been implemented in the MIDAS context by Andreou et al. (2013) using a very rich financial data set.

Note also that the direct multi-step forecasting strategy that we adopt provides parameter estimates $\beta^{(h)}$ and $\theta^{(h)}$ that depend on the prediction horizon h which is given at the lowest frequency time units. The selection parameters $\lambda_t^{(h)}$ for the LASSO approach and $\Omega_t^{(h)}$ for the Bayesian SSVS model are time dependent because their choice is based on the use of a recursive window framework over the whole out-of-sample period. Notice that since the selection variables ξ_i depend on $\lambda_t^{(h)}$ or $\Omega_t^{(h)}$ they are hence also time dependent. Note also that the FAMIDAS model and the forecast combination also involve time-varying specifications using the predictive cross-validation. Model settings are described below:

(i) In the case of the LASSO, we have the following forecasting equation:

$$\hat{y}_{t+h|t}(\lambda_t^{(h)}) = \hat{\beta}_0^{(h)} + \sum_{i=1}^{n} \xi_i(\lambda_t^{(h)}) \, \hat{\beta}_i^{(h)} \; m^{K_i}(\hat{\theta}_i^{(h)}, L) \, x_{t,i}^{\kappa_i}. \tag{LASSO-MIDAS}$$

As opposed to Tibshirani (1996), our goal is not to recover an underlying sparsity in the coefficients vector β but to use the penalty to reduce the covariates cardinality. The question that arises in this context is the selection of the optimal strength of the ℓ_1 penalty that ensures a favorable forecasting performance. Our cross-validating procedure follows those prescriptions: for a given value $\lambda > 0$, we set its corresponding selection by estimating the equation (4), and we forecast $\hat{y}_{t|t-h}(\lambda)$ such as defined in the LASSO-MIDAS equation. Then, repeating that for a range of λ, we determine $\lambda_t^\star$ as the one which minimizes forecasting residual at time t:

$$\lambda_t^{\star(h)} = \arg\min_{\lambda} \sum_{\mathsf{t}=t-d}^{t} \delta^{t-\mathsf{t}} \left(y_{\mathsf{t}} - \hat{y}_{\mathsf{t}|\mathsf{t}-h}(\lambda_{\mathsf{t}}^{(h)}) \right)^2, \tag{12}$$

where $\delta = 0.8$ to be coherent with our wish to involve a decrease on the MSFE weight with respect to the historical performance. That allows to promote recent forecasts accuracy.

(ii) The Bayesian Stochastic Search variable selection combined with the MIDAS forecasting model is given by:

$$\hat{y}_{t+h|t}(\Omega_t^{(h)}) = \hat{\beta}_0^{(h)} + \sum_{i=1}^{n} \xi_i(\Omega_t^{(h)})\, \hat{\beta}_i^{(h)}\; m^{K_i}(\hat{\theta}_i^{(h)}, L)\, x_{t,i}^{\kappa_i}. \tag{BAYESIAN-MIDAS}$$

The posterior probability ω_i as described in (9) and in (10) specifies the probability that β_i has not been draw from the spike prior, namely the probability to include it in the model. The issue that arises here is to choose a threshold $\Omega^\star \in [0,1]$ below which variables are simply removed. Following exactly the same procedure than in the LASSO case, we forecast $\hat{y}_{t|t-h}(\Omega)$ according to its relative set of selected variables. Then, we set $\Omega_t^\star$ as the minimum argument of the square error for the period t:

$$\Omega_t^{\star(h)} = \arg\min_{\Omega} \sum_{\mathsf{t}=t-d}^{t} \delta^{t-\mathsf{t}} \left(y_{\mathsf{t}} - \hat{y}_{\mathsf{t}|\mathsf{t}-h}(\Omega_t^{(h)}) \right)^2.$$

(iii) The FAMIDAS model minimize is basically based on a factor structure assumption for the explanatory variables matrix, That can be described as follows:

$$X_\tau = \Lambda F_\tau + \eta_\tau,$$

where τ is given in one of the higher frequencies (daily or monthly in our case). The components of the factors vector are denoted as $F_\tau = (f_{1,\tau}, \ldots, f_{r,\tau})$. This approach consists of using the standard MIDAS technique described in (1) with the r first estimated principal factors that are employed as explanatory variables. The model is given by

$$\hat{y}_{t+h|t}(r^{(h)}) = \hat{\beta}_0^{(h)} + \sum_{i=1}^{r^{(h)}} \hat{\beta}_i^{(h)} m^{K_i}(\hat{\theta}_i^{(h)}, L)\, \hat{f}_{t,i}^{\kappa_i} \tag{FAMIDAS}$$

Since factors are linearly uncorrelated, the size of the factor vector, r, can be determined with a statistical hypothesis test, like Bai and Ng (2008) have proposed. In our study, we propose to define $r^\star$ depending on the forecasting performances. In that case, the parameter we focus on is the number of factors to include in the final model.

$$r_t^{\star(h)} = \arg\min_r \sum_{\mathsf{t}=t-d}^{t} \delta^{t-\mathsf{t}} \left(y_\mathsf{t} - \hat{y}_{\mathsf{t}|\mathsf{t}-h}(r_t^{(h)}) \right)^2$$

Note that factors can only represent a family of variables sampled at the same frequency. Since we mix daily and monthly predictors, we define $r = (r^D, r^M)$, where $r^D = \{0, 1\}$ and $r^M = \{0, 1, 2\}$.

(iv) Combining forecasts is often referred as a good alternative to model selection. Formally, we compute n individual forecasts respectively based on the i^{th} variable of the entire set, as follows:

$$\hat{y}_{t+h|t,i} = \hat{\beta}_0^{(h)} + \hat{\beta}_i^{(h)}\; m^{K_i}(\hat{\theta}_i^{(h)}, L)\, x_{t,i}^{\kappa_i}. \tag{13}$$

The combination are then made using a weighted average of the individual forecasts (13), thus it can be written as follows:

$$\hat{y}_{t+h|t}(w_t^{(h)}) = \sum_{i=1}^{n} w_{t,i}^{(h)}\; \hat{y}_{t+h|t,i} \tag{COMBINATION}$$

The forecast relies on the vector of the time-varying combination weights $w_{i,t}^{(h)}$ which can be estimated using several methods; Stock and Watson (2008) show some of those techniques. In this paper, we determine using an equivalent procedure than others selection methods to fairly compare all models. This model relies on the vector of $w_{i,t}^\star$ that weights the individual forecasts, see (13). Those are given as follows:

$$w_{t,i}^{\star(h)} = \frac{\mu_{t,i}^{-a}}{\sum_{j=1}^{n} \mu_{t,j}^{-a}} \quad \text{where } \mu_{t,i} = \sum_{\mathsf{t}=t-d}^{t} \delta^{t-\mathsf{t}} \left(y_t - \hat{y}_{\mathsf{t}|\mathsf{t}-h,i} \right)^2, \text{ and } a = 2.$$

In these four models, the predictive cross-validation is based on forecasting performances over the d previous quarters. Notice that the value of d would have different meanings, e.g. $d = 1$ tells that we only base the analysis on the last period whereas $d = 20$ represents the selection that gave best results over the last 5 years. Furthermore, instead of the usual MSFE (Mean Squared Forecasting Errors), we prefer focusing on an discounted version of this criterion such as Andreou et al. (2013) used in their paper. That metric promotes recent performances by weighting squared residuals according to their historical records. As regards the pseudo out-of-sample period, we opt for a intermediate parametrization which corresponds to forecasting performances over the last year, i.e. $d = 4$.

Using this cross-validation procedure on previous quarters before the forecast stage $t + h$ within the recursive window framework that we describe above, the selection is updated every period of the out-of-sample. This technique lies in an automated model selection procedure that should lead to both better selection of the leading indicators and greater efficiency in their use.

6 Empirical exercise on US data

6.1 dataset

We assess the performance of the four models we have presented above using a forecasting exercise on US GDP data over the period 2000q1-2012q4 while the full sample covers a longer period going from 1964q3 to 2012q4. In this forecasting exercise, we focus on predicting the quarterly US Gross Domestic Product using a set of 24 variables which includes monthly real indicators and daily financial variables. More specially, the dataset incorporates a daily spread rate and three financial times series.Our set also includes seventeen monthly indicators representing the real US economy and coming from "soft" and "hard" data (production index, housing statistics, unemployment rate, opinion survey, etc.). An entire description of the dataset is available in Table 2 of Appendix C.

The severe recession has shed light on the necessary re-assessment of the contribution of financial markets to the economic cycles. There is a huge volume of work in the literature that underlines the leading role of financial variables in the forecasting of macroeconomic fluctuations. Recently, Chauvet et al. (2012) and Ferrara et al. (2014) have even shown that daily volatility of financial time series series have a significant forecasting power in explaining US growth. Using variable selection models within the predictive cross-validation we have put forward, we evaluate whether both returns and volatility of financial time series would be included in the model specifications to forecast US GDP growth. Given that volatility is not directly observable, several methods have been developed in the literature to estimate it. Following Ferrara et al. (2014), we use a GARCH model on whitened and winsorized daily financial series. Let ρ_τ denotes the daily returns of a given financial time series (τ corresponds to the daily frequency). The GARCH(1,1) specification is given by:

$$\begin{cases} \rho_\tau &= v_\tau \, \eta_\tau, \\ (v_\tau)^2 &= c + a\,(v_{\tau-1})^2 + b\,\rho^2_{\tau-1}, \end{cases} \tag{14}$$

where $\{\eta_\tau\} \sim \mathrm{WN}(0,1)$. In order to ensure the existence of a unique stationary solution and the positivity of the volatility, we assume that $a > 0$, $b \geqslant 0$, and $a+b < 1$. Estimated daily volatilities stemming from equation (14) are considered as explanatory variables of the US macroeconomic fluctuations.

6.2 Forecasting results

From 2000 to 2013, the US economy experiences different phases of the business cycle. In 2008, in the wake of the financial crisis, the United States entered a severe recession, referred to as the Great Recession. The recovery since 2009 was weak and growth remained uneven. Our approach allows to set the horizon at which leading indicators have early information and can send warnings of turning point. In this respect, we assess MIDAS-based models presented in the previous sections, by splitting our sample in three parts: Early 2000's (from 2000q1 to 2007q2), Great Recession (from 2007q3 to 2009q4), and Recovery (from 2010q1 to 2012q4). Table 1 reports the Mean Squared Forecasting Errors (MSFE) in these three periods. As a benchmark, we report results from a simple autoregressive model, AR(1), whose order has been determined by using Bayesian Information Criteria. Moreover, in order to assess the predictive gain of selecting variables, we also report results from a MIDAS model that makes use of the full set of variables. Point forecasts and model inclusion for all horizons are exhibited in the Appendix D. Results of the forecast comparison exercise for GDP growth are discussed below.

Nowcasting

At the short horizons, the indicators chosen by both variable selection techniques, the lasso-midas and the bayesian-midas, are primarily related to the real economic activity (production, labor market, housing, consumption). This stylized fact has been observed in empirical papers pointing out the increasing role of hard indicators on macroeconomic forecasts when we are close to the release date. In addition, at this horizon, the financial volatility of the SP500 was among the best predictors (always included in both predictor set).

Best performances in nowcasting the Great Recession were given by both variable selection methods. Those indicate that financial instability, especially observed via volatility variables and commodity price indices, triggered confusion and fear among consumers and firms. Lower confidence and lower stock price yield to a net decrease in consumption in the current period and hence in GDP growth. These findings are in agreement with the combination model in which IPI and the ISM PMI survey are particularly important during this period.

The 3-month to 6-month ahead horizons

At the 6-month-ahead horizon, financial variables emerge as the most useful indicators. Spread rate and stock price volatility dominate the top ranks in model inclusion. A key difference between pure nowcasting at 0-month-ahead and 3-month-ahead forecasts is that at the latter horizon, IPI variables are not very prominent. This result is interesting for practitioners in the sense that using industrial production index at this horizon does not appear useful. We also note that the lasso-midas tends to select variables that are not encompassed by other indicators. In fact, the LASSO would prefer substitution in spite of complementarity that could be involved by the bayesian-midas shrinkage and the forecast combination.

	2000q1-2012q4 Full sample	2000q1-07q2 Early 2000's	2007q3-09q4 Great Recession	2010q1-12q4 Recovery
$h = 0$ (Nowcasting)				
LASSO-MIDAS	0,34	0,33	0,45	0,19
BAYESIAN-MIDAS	0,32	0,29	0,52	0,20
FAMIDAS	0,33	0,28	0,73	0,14
COMBINATION	0,38	0,33	0,62	0,12
MIDAS	0,43	0,52	0,54	0,23
AR	0,44	0,30	1,16	0,16
$h = 3$				
LASSO-MIDAS	0,37	0,32	0,79	0,13
BAYESIAN-MIDAS	0,40	0,37	0,79	0,15
FAMIDAS	0,40	0,27	1,14	0,12
COMBINATION	0,42	0,33	0,94	0,20
MIDAS	0,51	0,55	0,84	0,23
AR	0,47	0,28	1,47	0,17
$h = 6$				
LASSO-MIDAS	0,52	0,34	1,24	0,18
BAYESIAN-MIDAS	0,48	0,37	1,14	0,19
FAMIDAS	0,46	0,30	1,39	0,13
COMBINATION	0,42	0,30	0,99	0,27
MIDAS	0,62	0,49	1,49	0,29
AR	0,55	0,32	1,74	0,17
$h = 9$				
LASSO-MIDAS	0,55	0,35	1,53	0,23
BAYESIAN-MIDAS	0,47	0,31	1,19	0,21
FAMIDAS	0,52	0,33	1,65	0,15
COMBINATION	0,42	0,32	1,07	0,16
MIDAS	0,58	0,52	1,45	0,29
AR	0,55	0,34	1,87	0,19
$h = 12$				
LASSO-MIDAS	0,54	0,31	1,61	0,22
BAYESIAN-MIDAS	0,66	0,38	2,10	0,18
FAMIDAS	0,52	0,29	1,76	0,15
COMBINATION	0,44	0,30	1,12	0,25
MIDAS	0,68	0,46	1,85	0,26
AR	0,58	0,32	1,99	0,19

Table 1: MSFE (Mean Squared Forecasting Errors)

Regarding performances in predicting the Great Recession, these four models have captured early warnings from 6 to 3 months ahead. By the end of 2007, serious short-term risks were looming: uncertainty on financial markets (captured by stocks price volatility), bank loan contraction, rising interest rates. We note that model inclusion of those indicators in both variable selection and an increasing weight in the combination. From early 2010 to the end of 2011, while financial indices remain high, the recovery was slower than expected, it was referred sometimes in the literature to as *Sluggish Recovery*. Both FAMIDAS and COMBINATION models show this disconnection by either not including anymore the daily factor or by reducing their weights in the regression. Finally, by the end of 2011, in the wake of the sovereign debt crisis, some financial indicators (spread rate, corporate bonds and stocks volatility) were again chosen.

The 9-month to 12-month ahead horizons

At the 9-month ahead horizon, we note that financial volatility indicators already played an important role in forecasting, especially over the Early 2000's and the Recovery, as already noticed for the 6-month-ahead forecasts. In addition at this horizon, we get complementary information from money related variables such currency component of M1, M2 and St Louis monetary base. We also find that inflation rate (CPI) is chosen by both variable selection and highlighted as one the main indicator in the combination model, at both 9-month and 12-month ahead horizons. Findings during the Great Recession period should be interpreted with care since forecasting errors are really high. Indeed, from 12 to 9 months ahead, it turns out that the four models provided flat predictions, and hence did not yield informative contents to anticipate the crisis.

Summary

According to our results of the forecast comparison exercise for GDP growth, four main conclusions can be drawn. First, we note that over the whole sample, our four MIDAS based models outperforms both the autoregressive and the full MIDAS models, and we also notice that forecasting errors for all models decrease when the forecasting horizon tends to zero. Second, results significantly differ depending on the period. More specifically, we observe that early warnings of the "Great Recession" were really identify from 3 to 6 months ahead. In fact, models we have studied in general tend to perform best with short horizons although in some cases the performance extends to three or four quarters. Third, a few economic stylized facts have been summarized as regards the set of predictors and the forecasting horizon. The set of chosen indicators includes reasonable variables from an economic point of view and reflects both their intrinsic leading features and the time varying nature of current economic outlook. Fourth, our forecasting exercise on GDP growth proves that pooling indicators ability provides very reliable models. Observed individually over their respective primary horizons, some of indicators would already give very good results, grouping yields even better performances and minimum forecast errors.

7 Conclusion

This paper considers variable selection in short-term macroeconomic forecasting. From the perspective of prediction, we ground our analysis on a mixed frequency framework, the MIDAS, that allows the use of any available data regardless of their sampling frequency.

We specially develop in this context, four tools to identify leading indicators of the US GDP growth using an automatic model selection procedure based on recent best performances. More specifically, we introduce a LASSO augmented MIDAS model and a Bayesian MIDAS Stochastic Search Variable Selection that we compare with the Factor Augmented MIDAS model, and the combination forecast technique of univariate MIDAS models. Those are combined with a predictive cross-validation methodology which uses a recursive window and specifies the set of predictors with respect to their ability. Dimension reduction methods that we use, goes beyond point forecast and highlights the leading role of some indicators in macroeconomics. Our findings emphasize the role of daily financial information in predicting GDP and show that combining daily and monthly indicators increases the forecasting accuracy.

The generic question we addressed focuses on variable selection in predictive mixed frequency models. Forecasting GDP is only one of many examples where our methods can be applied. Our methods have broad applicability and can be of general interest in many other macroeconomic applications.

References

Andreou, E., Ghysels, E., and Kourtellos, A. (2010). Regression models with mixed sampling frequencies. *Journal of Econometrics*, 158(2):246–261.

Andreou, E., Ghysels, E., and Kourtellos, A. (2013). Should macroeconomic forecasters use daily financial data and how? *Journal of Business and Economic Statistics*, 31(2):240–251.

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

Banbura, M., Giannone, D., Modugno, M., and Reichlin, L. (2012). Now-Casting and the Real-Time Data Flow. Technical Report 9112, CEPR.

Barhoumi, K., Benk, S., Cristadoro, R., Reijer, A. D., Jakaitiene, A., Jelonek, P., Rua, A., Rünstler, G., Ruth, K., Nieuwenhuyze, C. V., and Jakaitiene, A. (2008). Short-term forecasting of GDP using large monthly datasets: a pseudo real-time forecast evaluation exercise. Working paper, European Central Bank.

Barhoumi, K., Darné, O., and Ferrara, L. (2010). Are disaggregate data useful for factor analysis in forecasting French GDP? *Journal of Forecasting*, 29(1-2):132–144.

Bencivelli, L., Marcellino, M., and Moretti, G. (2012). Selecting predictors by using Bayesian model averaging in bridge models.

Chauvet, M., Senyuz, Z., and Yoldas, E. (2012). What does financial volatility tell us about macroeconomic fluctuations? Working paper, Federal Reserve Board.

Clements, M. P. and Galvão, A. B. (2009). Forecasting US output growth using Leading Indicators: An appraisal using MIDAS models. *Journal of Applied Econometrics*, 24(7):1187–1206.

De Mol, C., Giannone, D., and Reichlin, L. (2008). Forecasting using a large number of predictors: Is Bayesian shrinkage a valid alternative to principal components? *Journal of Econometrics*, 146(2):318–328.

Efron, B., Hastie, T., Johnstone, I., and Tibshirani, R. (2004). Least angle regression. *Annals of Statistics*, 32(2):407–499.

Ferrara, L. and Marsilli, C. (2013). Financial variables as leading indicators of GDP growth: Evidence from a MIDAS approach during the Great Recession. *Applied Economics Letters*, 20(3):233–237.

Ferrara, L. and Marsilli, C. (2014). Nowcasting global economic growth: A factor-augmented mixed-frequency approach. Technical Report 515, Banque de France.

Ferrara, L., Marsilli, C., and Ortega, J.-P. (2014). Forecasting growth during the Great Recession: is financial volatility the missing ingredient? *Economic Modelling*, 36:44–50.

Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2000). The Generalized Dynamic-Factor Model: Identification And Estimation. *The Review of Economics and Statistics*, 82(4):540–554.

Foroni, C. and Marcellino, M. (2014). A comparison of mixed approaches for modelling euro area macroeconomic variables. *International Journal of Forecasting*, 30(3):554–568.

Foroni, C., Marcellino, M., and Schumacher, C. (2013). U-MIDAS: MIDAS regressions with unrestricted lag polynomials. *Journal of the Royal Statistical Society - Series A*, forthcomin.

George, E. I. and McCulloch, R. E. (1993). Variable Selection Via Gibbs Sampling. *Journal of the American Statistical Association*, 88(423):881–889.

George, E. I. and McCulloch, R. E. (1997). Approaches for Bayesian variable selection. *Statistica Sinica*, 7:339–373.

Ghysels, E. (2012). Mixed Frequency Vector Autoregressive Models.

Ghysels, E., Santa-clara, P., and Valkanov, R. (2002). The MIDAS Touch: Mixed Data Sampling Regression Models.

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

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

Ishwaran, H. and Rao, J. S. (2005). Spike and slab variable selection: Frequentist and Bayesian strategies. *The Annals of Statistics*, 33(2):730–773.

Kaufmann, S. and Schumacher, C. (2012). Finding relevant variables in sparse Bayesian factor models: Economic applications and simulation results. Technical Report 29, Deutsche Bundesbank Discussion Paper.

Korobilis, D. (2013). Hierarchical shrinkage priors for dynamic regressions with many predictors. *International Journal of Forecasting*, 29(1):43–59.

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

Malsiner-Walli, G. and Wagner, H. (2011). Comparing Spike and Slab Priors for Bayesian Variable Selection. *Austrian Journal of Statistics*, 40(4):241–264.

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

Mitchell, T. J. and Beauchamp, J. J. (1988). Bayesian Variable Selection in Linear Regression. *Journal of the American Statistical Association*, 83(404):1023–1032.

Nesterov, Y. (2005). Smooth minimization of non-smooth functions. *Math. Program.*, 103(1, Ser. A):127–152.

Ng, S. (2012). Variable Selection in Predictive Regressions.

Palm, F. C. and Zellner, A. (1992). To combine or not to combine? issues of combining forecasts. *Journal of Forecasting*, 11(8):687–701.

Rodriguez, A. and Puggioni, G. (2010). Mixed frequency models: Bayesian approaches to estimation and prediction. *International Journal of Forecasting*, 26(2):293–311.

Schumacher, C. (2010). Factor forecasting using international targeted predictors: The case of German GDP. *Economics Letters*, 107(2):95–98.

Scott, S. L. and Varian, H. R. (2013). Bayesian variable selection for nowcasting economic time series. Technical Report 19567, NBER Working Papers.

Stock, J. H. and Watson, M. W. (2002). Forecasting Using Principal Components From a Large Number of Predictors. *Journal of the American Statistical Association*, 97:1167–1179.

Stock, J. H. and Watson, M. W. (2008). Phillips Curve Inflation Forecasts. Technical Report 14322, NBER Working Papers.

Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society*, 58(1):267–288.

Timmermann, A. (2006). Forecast combinations. In Elliott, G., Granger, C. W., and Timmermann, A., editors, *Handbook of Economic Forecasting*, chapter 4. Elsevier.

Yuan, M. and Lin, Y. (2005). Efficient Empirical Bayes Variable Selection and Estimation in Linear Models. *Journal of the American Statistical Association*, 100(472):1215–1225.

Appendices

A Nesterov regularization technique

Let us consider the following regression model:

$$\widehat{\beta} = \arg\min_{\beta} \sum_{t} \left(y_t - \beta_0 - \sum_{i} \beta_i \, x_{t,i} \right)^2 + \lambda_{\text{lasso}} \sum_{i} |\beta_i| \tag{15}$$

where y_t is the dependent variable, x_t is the vector of covariates, β is the vector containing the regression parameters, and λ_{lasso} is the exogenous parameter which controls the strength of the LASSO sparsifying regularization. To overcome the estimation of the problem 15, we use the following local regularization technique of Nesterov (2005).

We start by noting that the ℓ_1 norm can be expressed using the function g defined as

$$g(\beta) = \|\beta\|_1 = \max_{\|\gamma\|_\infty \leqslant 1} \gamma' \beta.$$

Then, we define the function g_μ such that $g_\mu \longrightarrow g$ with respect to $\mu \to 0$ and $\mu > 0$. We have:

$$g_\mu(\beta) := \max_{\|\gamma\|_\infty \leqslant 1} \gamma' \beta - \frac{\mu}{2} \|\gamma\|_2^2,$$

The Nesterov regularization technique consists of replacing the norm $g(\beta) = \|\beta\|_1$ by $g_\mu(\beta)$ with μ small. The advantage of proceeding in this fashion is that the function g_μ is obviously smooth with a gradient $\nabla g_\mu(\beta)$ whose components are given by

$$\nabla_i g_\mu(\beta) = \begin{cases} \text{sign}(\beta_i) & \text{if } |\beta_i| > \mu, \\ \frac{1}{\mu} \beta_i & \text{if } |\beta_i| < \mu. \end{cases}$$

B MCMC algorithm

To estimate the SSVS-MIDAS model, we implement a Gibbs sampler with respect to feature specifications. The algorithm relies on a few steps which successively sampling h from the spike and slab prior, the hyperparameter ω from a Beta distribution, β and σ from the usual Normal-Inverse Gamma prior, and θ from a candidate generating density using an Independence chain Metropolis-Hastings algorithm. Given initial values for all unknown parameters, the algorithm iteratively updates their values by sampling from their conditional distribution and hence constructing a Markov chain with an invariant distribution.

The algorithm is constructed as follows:

1. Sample $h_i,\ \forall i = 1, \ldots, n$,
   $\pi(h_i|\beta_i, \omega_i) = (1-\omega)\pi(\beta_i; 0, c\varphi^2)I_{\{h_i=c\}} + \omega\pi(\beta_i; 0, \varphi^2)I_{\{h_i=1\}}$,

2. Sample ω from $\mathcal{B}(c_0 + n_1, d_0 + n - n_1)$,
   where $n_1 = \sum_i \mathrm{I}_{\{h_i=1\}}$

3. Sample $\beta_i \sim \mathcal{N}(a_n, A_n)$
   where $A_n^{-1} = \frac{1}{\sigma}\mathsf{X}(\theta)'\mathsf{X}(\theta) + D^{-1}$, $a_n = A_n\frac{\mathsf{X}(\theta)Y}{\sigma}$, and $D = \mathrm{diag}(\phi^2 h_i)$

4. Sample $\sigma \sim \mathcal{IG}(s_n, S_n)$
   where $s_n = s_0 + \frac{T-1}{2}$, and $S_n = \frac{1}{2}(Y - \mathsf{X}(\theta)\beta)'(Y - \mathsf{X}(\theta)\beta)$

5. Sampling θ using an independence chain Metropolis-Hasting algorithm. The acceptance probability α to change to the new value θ^{new} drawn from the candidate density, determines whether the chain moves from areas of low posterior probability to high. It can be written as:

$$\alpha = \min\left[\frac{\pi(\theta = \theta^{\text{new}}|y)}{\iota(\theta = \theta^{\text{new}})}\frac{\iota(\theta = \theta^{\text{old}})}{\pi(\theta = \theta^{\text{old}}|y)}, 1\right].$$

   To define the candidate generating density ι, we use an approximation based on the asymptotic normality of the maximum likelihood estimator $\hat{\theta}_{ML}$, and on its asymptotic variance-covariance matrix $var(\hat{\theta}_{ML}) = \mathcal{I}(\theta)^{-1}$. We compute the Fisher information matrix $\mathcal{I}(\theta) = -\mathrm{E}\left(\frac{\partial^2}{\partial\theta\partial\theta'}\log f(Y|\beta, \theta, \sigma)\right)$, using numerical differentiation procedures to obtain the approximate variance: $\widehat{var(\hat{\theta}_{ML})}$. Thus, we set the candidate generating density as $\iota(\theta) = f_T(\theta|\hat{\theta}_{ML}, \widehat{var(\hat{\theta}_{ML})})$ since we approximate the posterior by a multivariate normal distribution with mean $\hat{\theta}_{ML}$ and covariance matrix $\widehat{var(\hat{\theta}_{ML})}$.

   Draw $u \sim \mathcal{U}(0, 1)$. If $u < \alpha$, retain the new candidate θ^{new} by setting $\theta = \theta^{\text{new}}$, otherwise $\theta = \theta^{\text{old}}$.

Repeating 25000 times these 5 steps yields the chain to converge to a steady state. The posterior distribution allows us to determine the selection with respect to ω. The MATLAB code will be available soon in my website: www.seltenhut.com/clement.marsilli.

C Data set

Daily series		
10y-3m	Spread rate: 10y Treasury Rate - 3m Treasury Bill	daily Δ
CRB	CRB Spot index, commodities price index	daily Δ log
DJ	Dow Jones industrial share price index	daily Δ log
SP500	S&P500 index	daily Δ log
CRBvolat	CRB Spot index, commodities price index	daily volatility (see 14)
DJvolat	Dow Jones industrial share price index	daily volatility (see 14)
SP500volat	S&P500 index	daily volatility (see 14)
Monthly series		
AAA	Moody Yield on Seasoned Corporate Bonds AAA	monthly Δ log
AMBSL	St Louis Adjusted Monetary Base	monthly Δ log
BAA	Moody Yield on Seasoned Corporate Bonds BAA	monthly Δ log
BusLoans	Commercial and Industrial Loans at Commercial banks	monthly Δ log
CPI	Consumer Price Index for all Urban Consumers: All items	monthly Δ log
Curr	Currency component of M1	monthly Δ log
DSPIC	Real Disposable Personal Income	monthly Δ log
Housing	New privately owned housing units started	monthly Δ log
IPI	Industrial Production Index	monthly Δ log
Loans	Loans and leases in bank credit, all commercial banks	monthly Δ log
M2	M2 money stock	monthly Δ log
Oil	Spot oil price: WTI	monthly Δ log
PCE	Personal Consumption Expenditures	monthly Δ log
PMI	ISM manufacturing survey: PMI composite index	monthly level
PPI	Producer Price Index: all commodities	monthly Δ log
TotalSL	Total consumer credit owned and securitized outstanding	monthly Δ log
Unemploy.	Unemployment rate	monthly Δ

Table 2: US data set from 1964:1 to 2012:4

D Empirical results

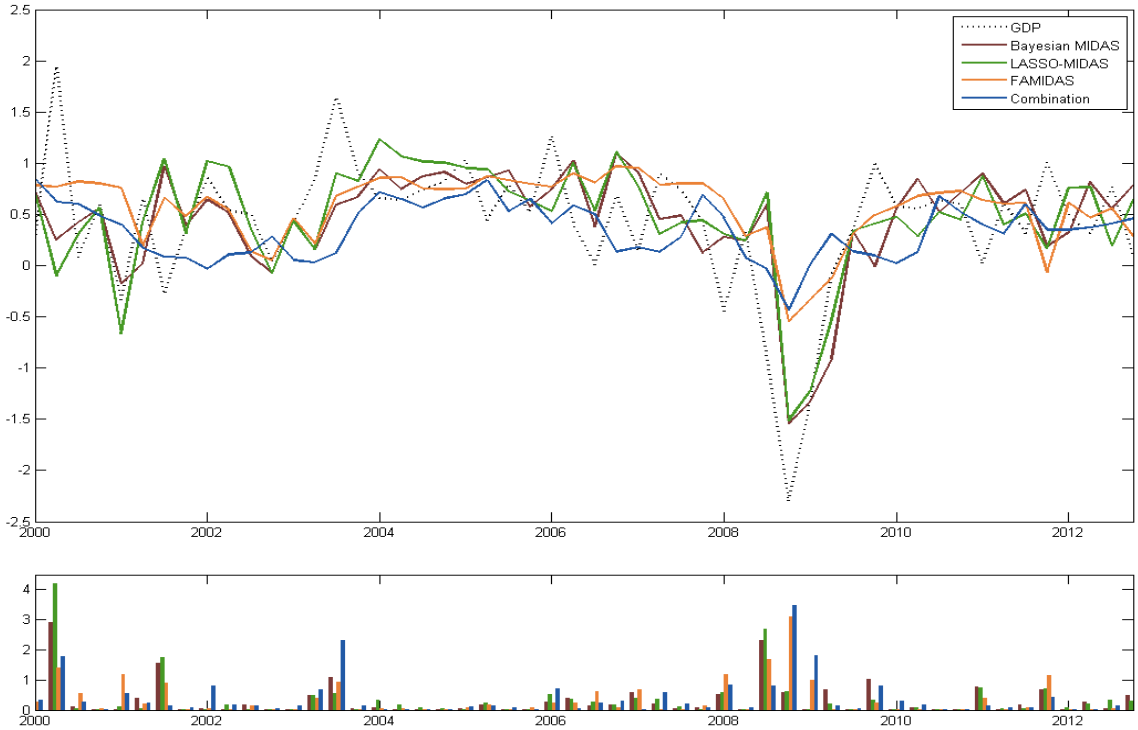

FIGURE 3: Point forecasts (top) and squared errors (bottom) for $h = 0$

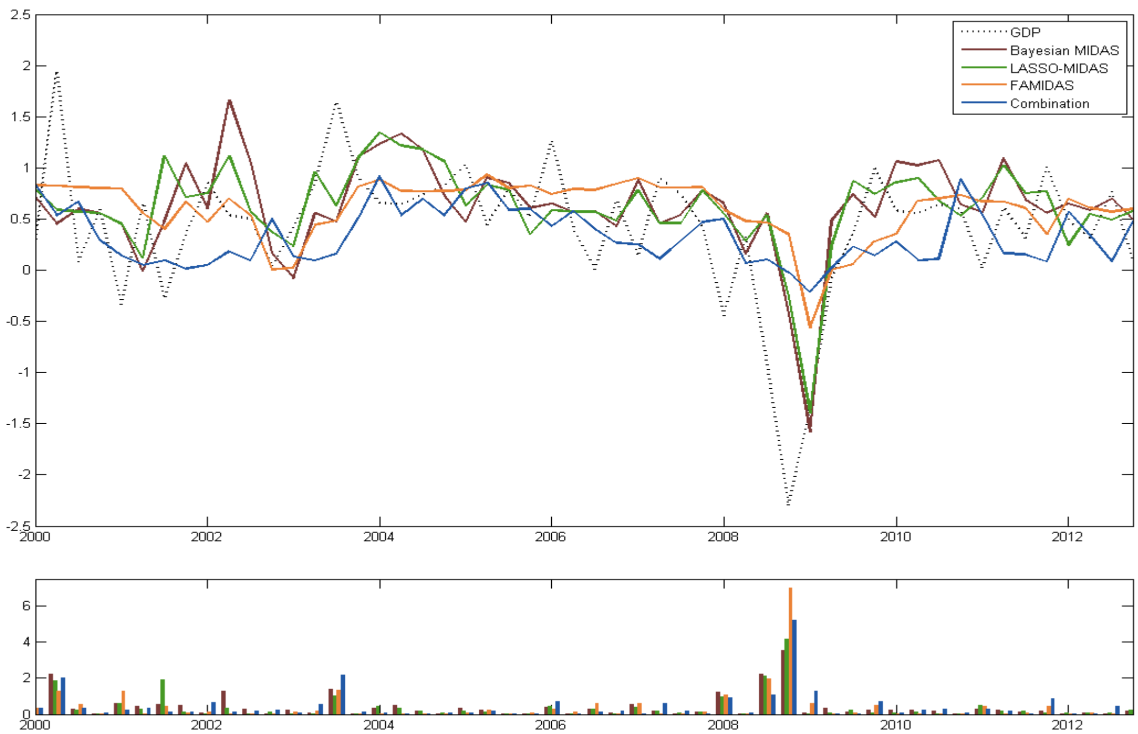

FIGURE 4: Point forecasts (top) and squared errors (bottom) for $h = 3$

D.1 Results for $h = 0$

Bayesian-MIDAS ($h = 0$)

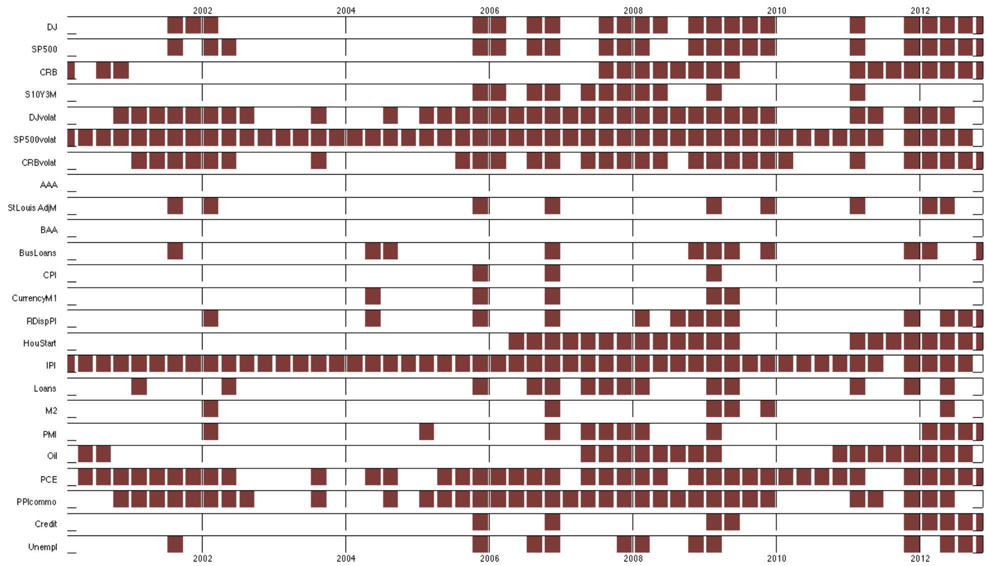

Figure 5: Variable selection from 2000q1 to 2012 q4 with the Bayesian-MIDAS model

LASSO-MIDAS ($h = 0$)

Figure 6: Variable selection from 2000q1 to 2012 q4 with the LASSO-MIDAS model

FAMIDAS ($h = 0$)

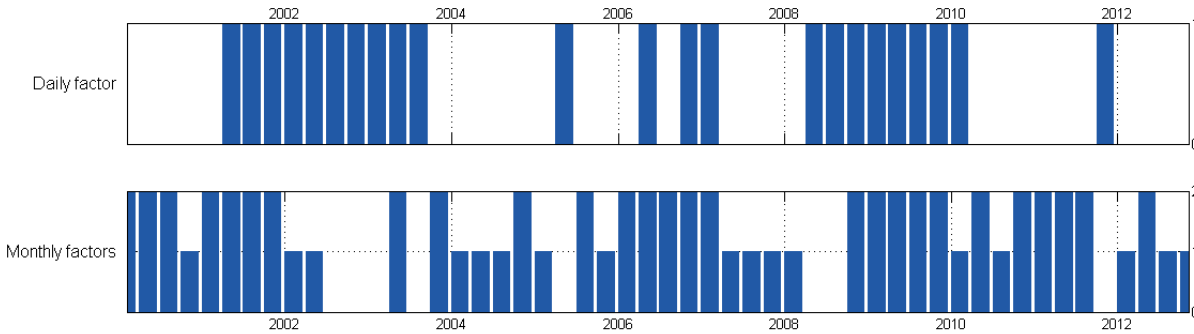

FIGURE 7: Variable selection from 2000q1 to 2012 q4 with the FAMIDAS model

Forecast combinations ($h = 0$)

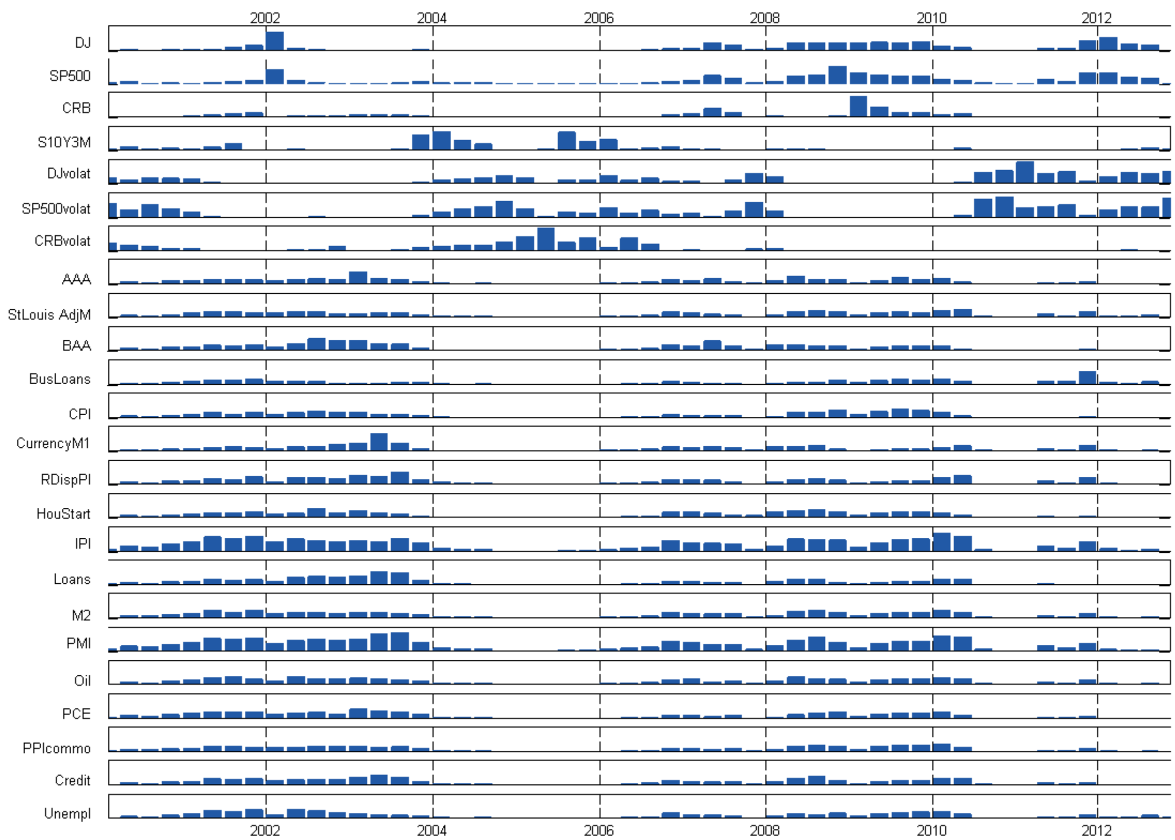

FIGURE 8: Weights for each variable of the combination from 2000q1 to 2012 q4

D.2 Results for $h = 3$

Bayesian-MIDAS ($h = 3$)

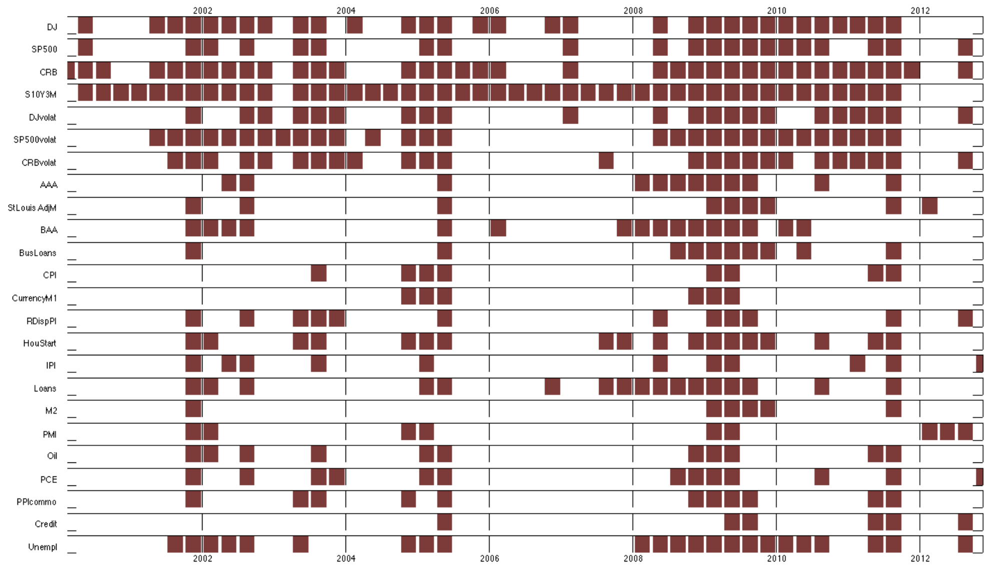

Figure 9: Variable selection from 2000q1 to 2012 q4 with the Bayesian-MIDAS model

LASSO-MIDAS ($h = 3$)

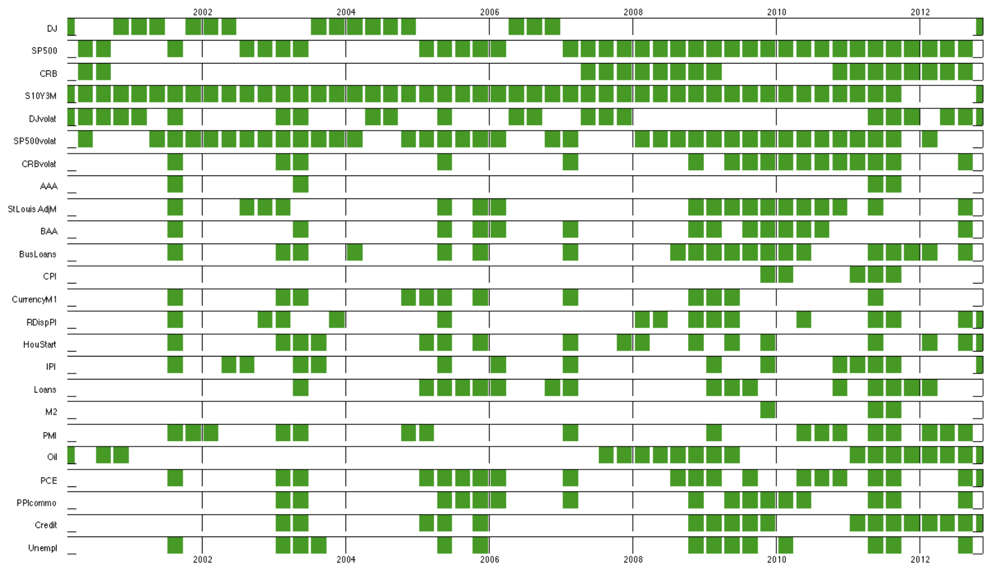

Figure 10: Variable selection from 2000q1 to 2012 q4 with the LASSO-MIDAS model

FAMIDAS ($h = 3$)

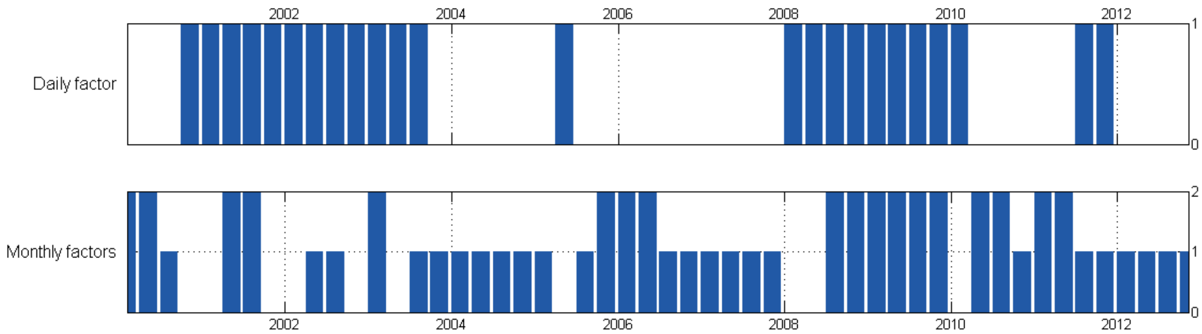

FIGURE 11: Variable selection from 2000q1 to 2012 q4 with the FAMIDAS model

Forecast Combinations ($h = 3$)

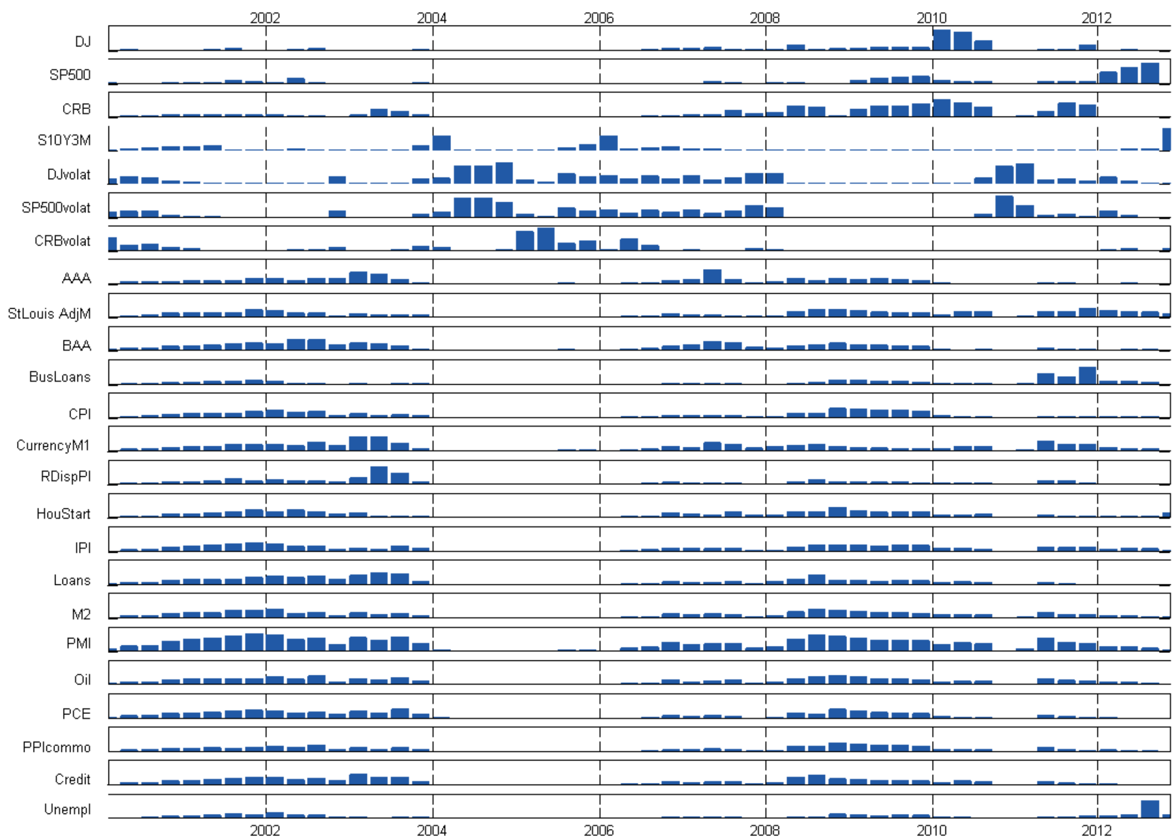

FIGURE 12: Weights for each variable of the combination from 2000q1 to 2012 q4

D.3 Results for $h = 6$

Bayesian-MIDAS ($h = 6$)

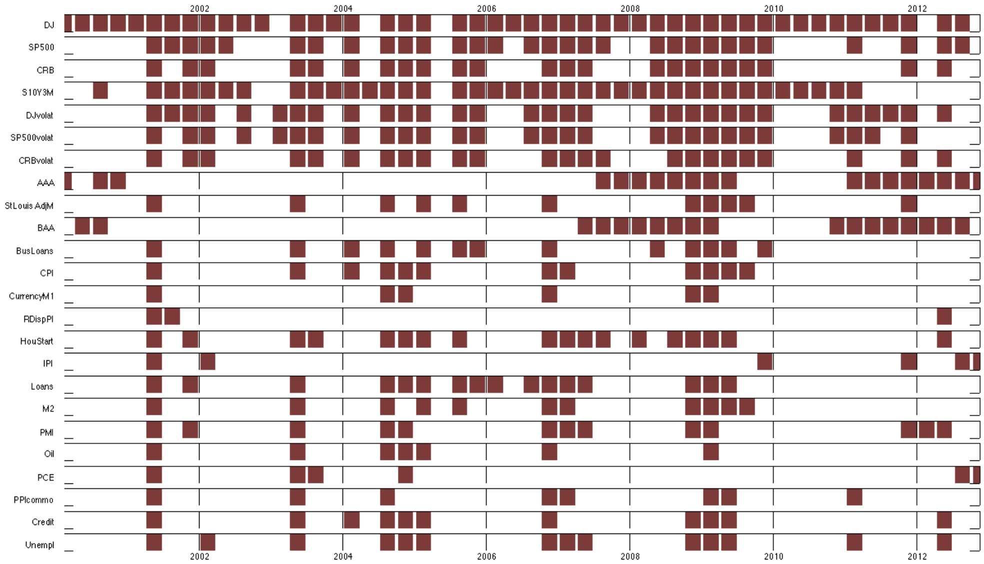

FIGURE 13: Variable selection from 2000q1 to 2012 q4 with the Bayesian-MIDAS model

LASSO-MIDAS ($h = 6$)

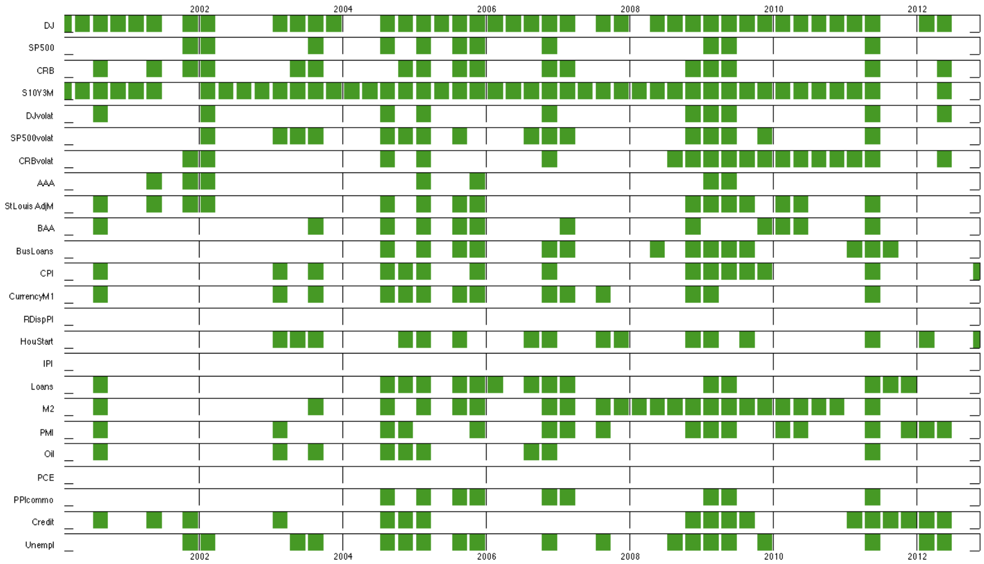

FIGURE 14: Variable selection from 2000q1 to 2012 q4 with the LASSO-MIDAS model

FAMIDAS ($h = 6$)

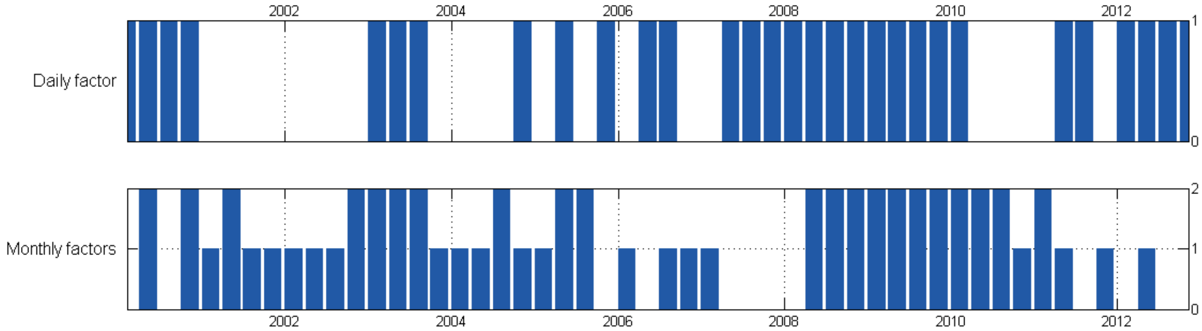

Figure 15: Variable selection from 2000q1 to 2012 q4 with the FAMIDAS model

Forecast combinations ($h = 6$)

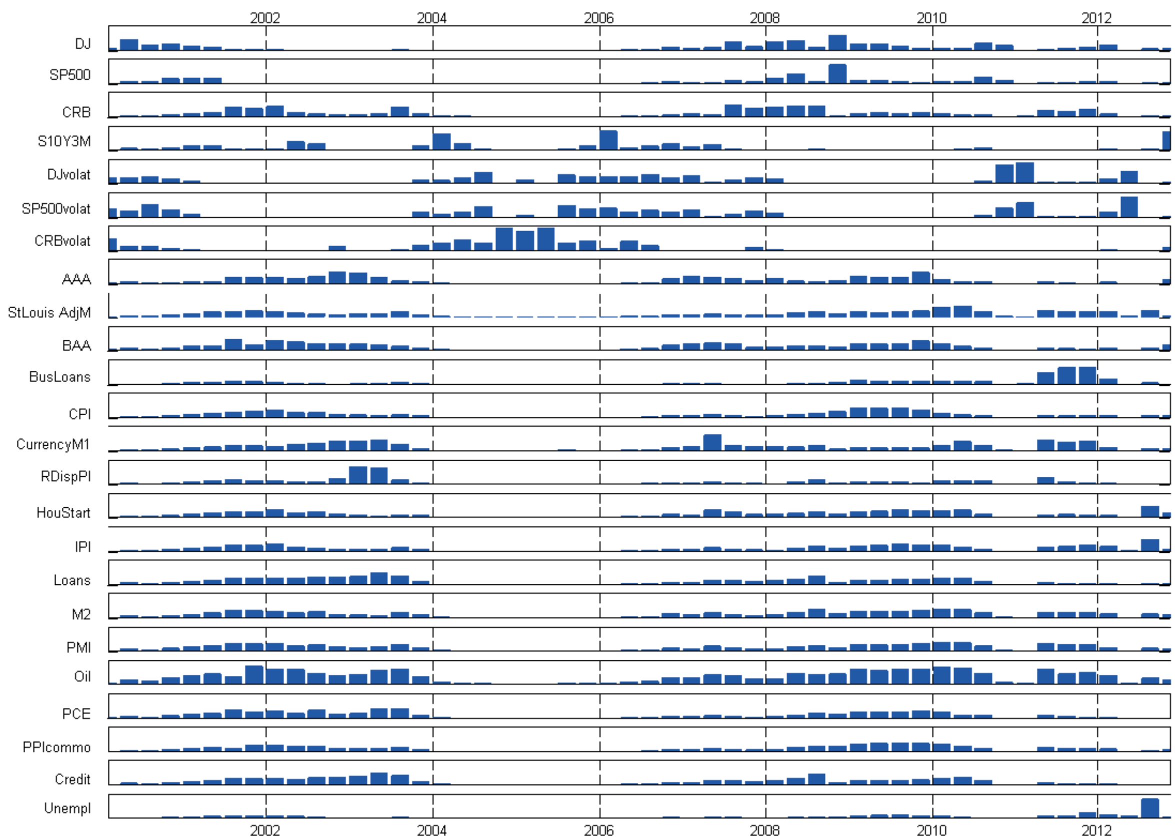

Figure 16: Weights for each variable of the combination from 2000q1 to 2012 q4

D.4 Results for $h = 9$

Bayesian-MIDAS ($h = 9$)

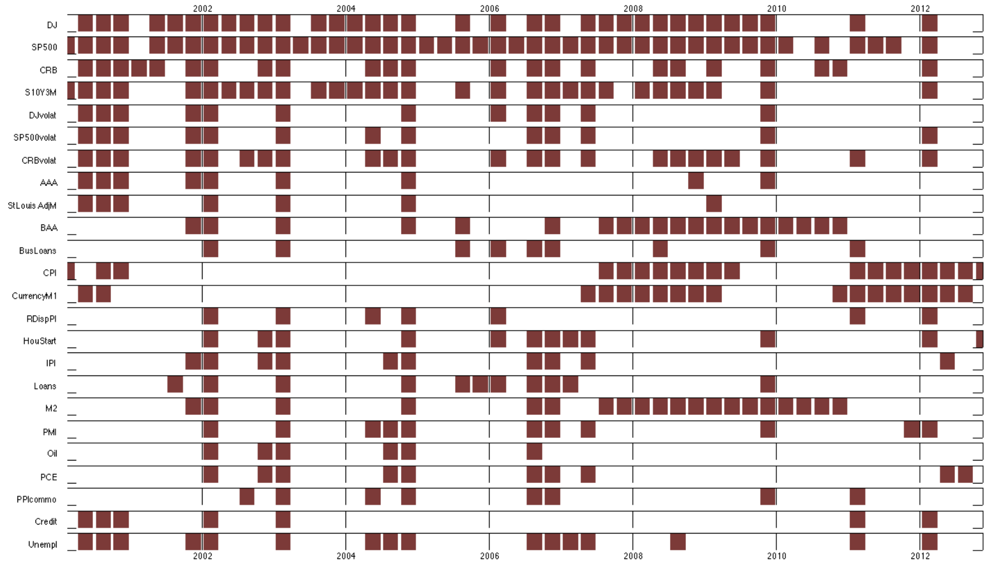

Figure 17: Variable selection from 2000q1 to 2012 q4 with the Bayesian-MIDAS model

LASSO-MIDAS ($h = 9$)

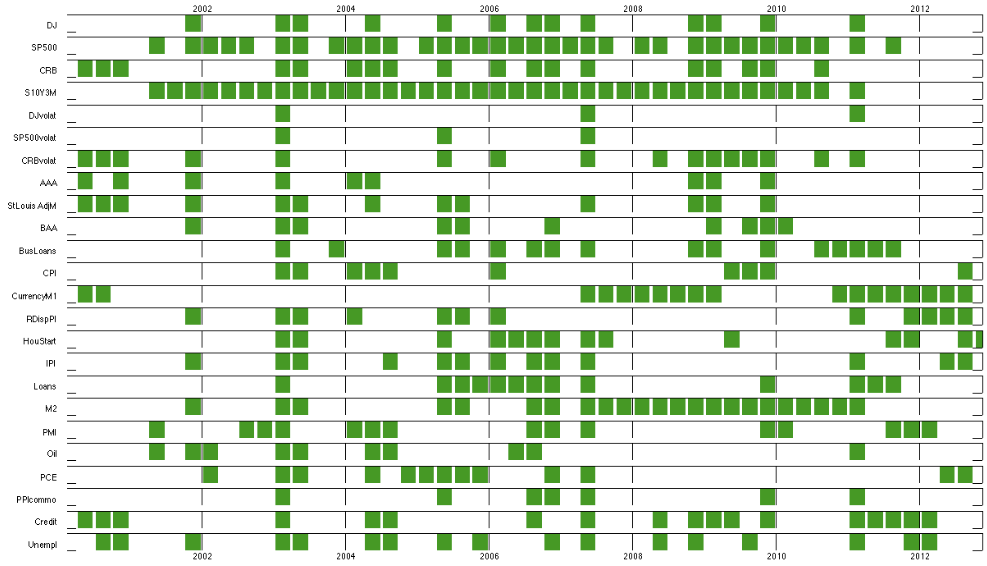

Figure 18: Variable selection from 2000q1 to 2012 q4 with the LASSO-MIDAS model

FAMIDAS ($h = 9$)

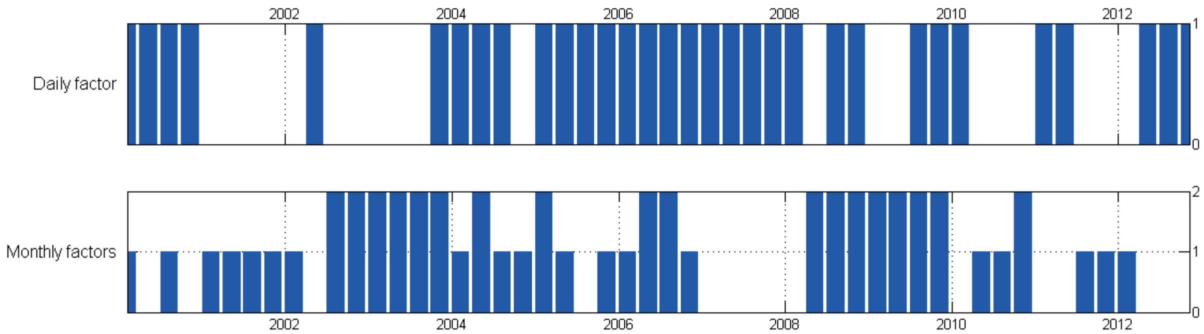

FIGURE 19: Variable selection from 2000q1 to 2012 q4 with the FAMIDAS model

Forecast combinations ($h = 9$)

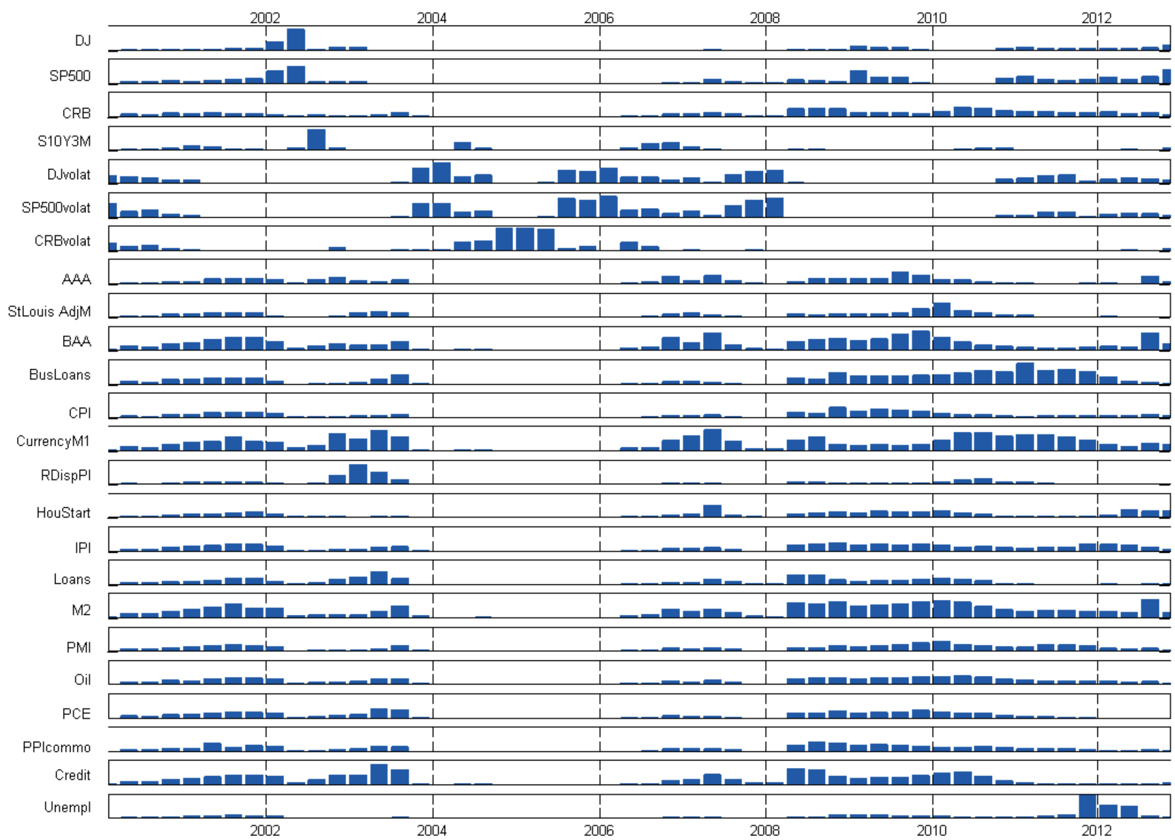

FIGURE 20: Weights for each variable of the combination from 2000q1 to 2012 q4

D.5 Results for $h = 12$

Bayesian-MIDAS ($h = 12$)

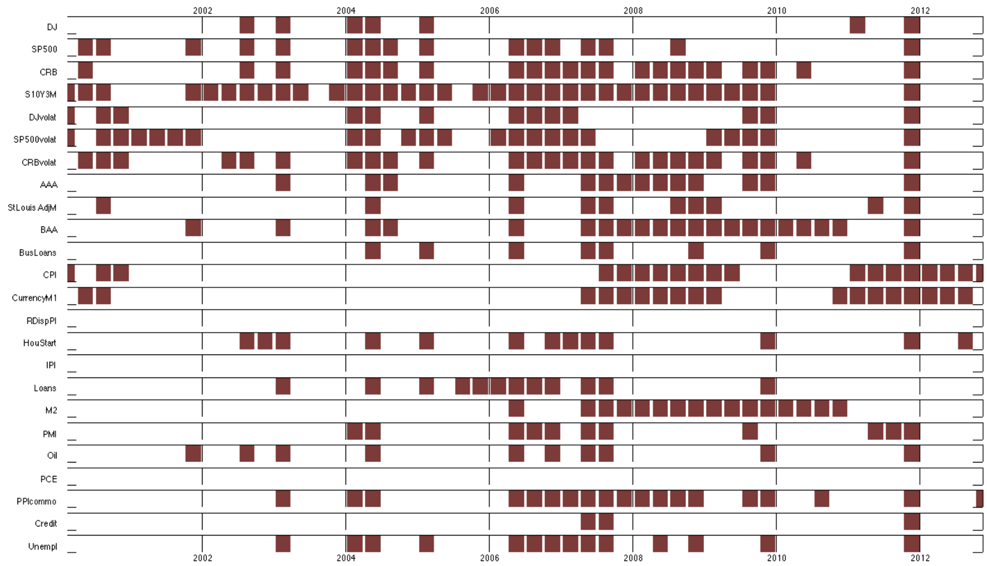

Figure 21: Variable selection from 2000q1 to 2012 q4 with the Bayesian-MIDAS model

LASSO-MIDAS ($h = 12$)

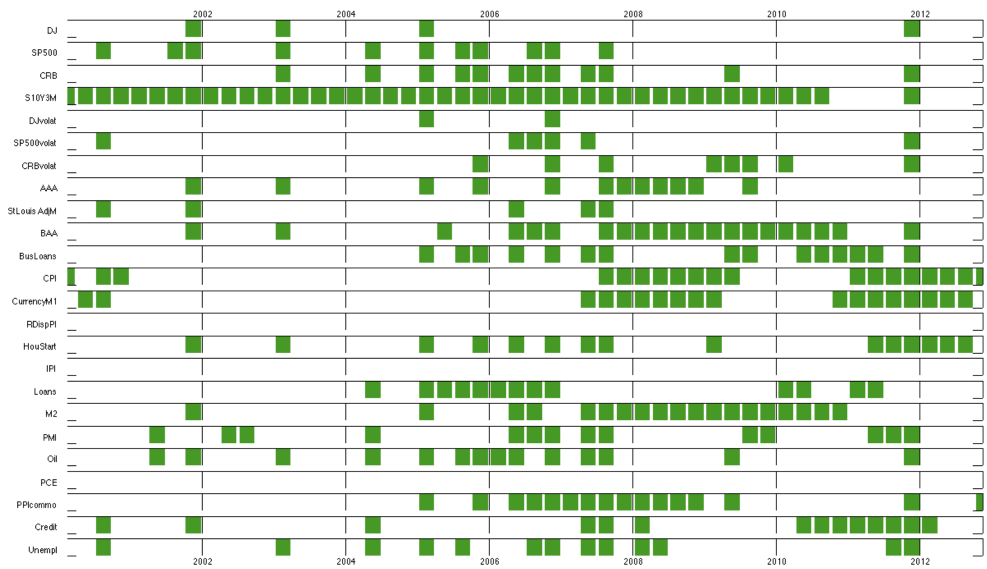

Figure 22: Variable selection from 2000q1 to 2012 q4 with the LASSO-MIDAS model

FAMIDAS ($h = 12$)

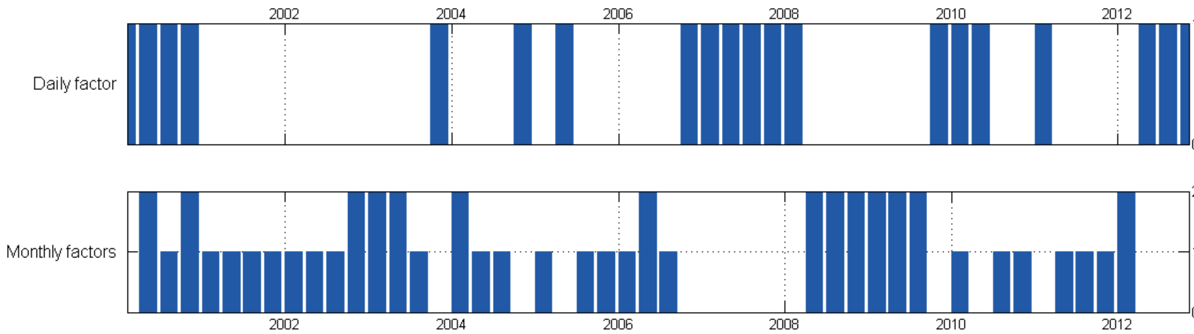

FIGURE 23: Variable selection from 2000q1 to 2012 q4 with the FAMIDAS model

Forecast combinations ($h = 12$)

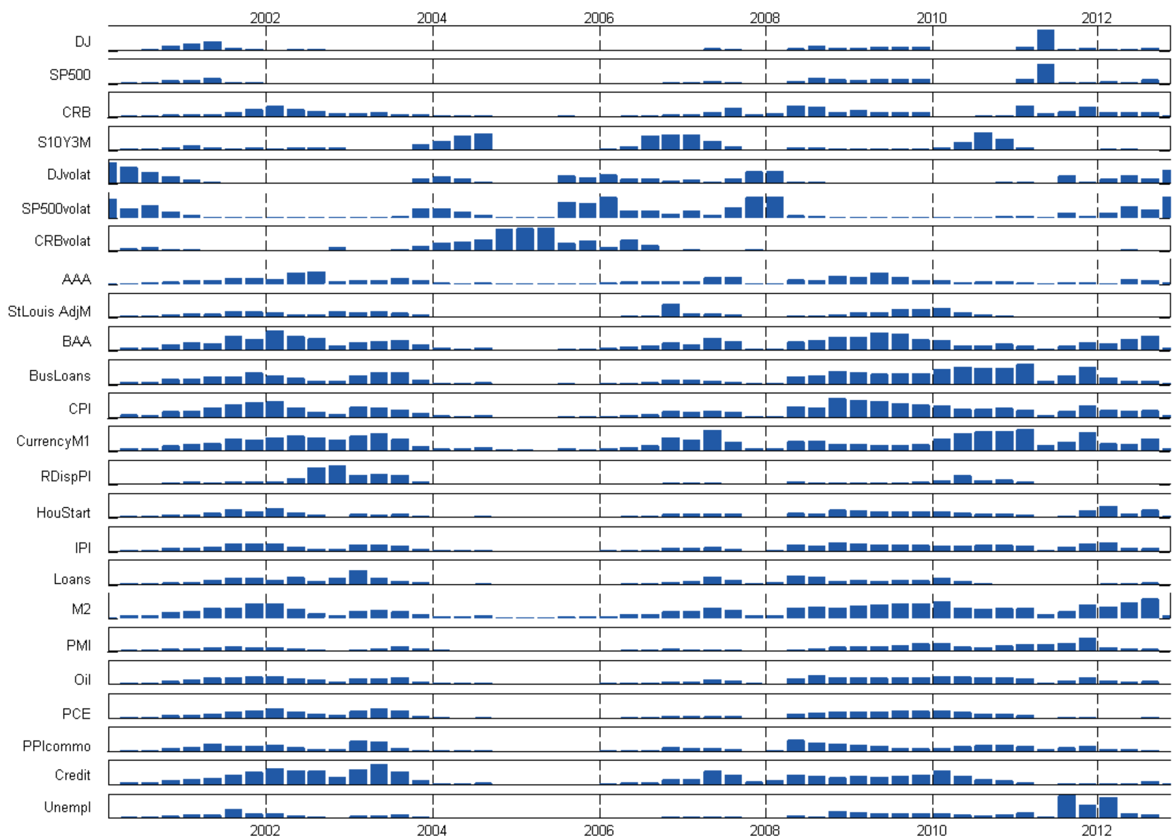

FIGURE 24: Weights for each variable of the combination from 2000q1 to 2012 q4

Documents de Travail

500. L. Behaghel, D. Blanche, and M. Roger, “Retirement, early retirement and disability: explaining labor force participation after 55 in France,” July 2014

501. D. Siena, “The European Monetary Union and Imbalances: Is it an Anticipation Story?,” July 2014

502. E. Mengus, “International Bailouts: Why Did Banks' Collective Bet Lead Europe to Rescue Greece?,” August 2014

503. J. Baron and J. Schmidt, “Technological Standardization, Endogenous Productivity and Transitory Dynamics,” August 2014

504. L. Arrondel, L. Bartiloro, P. Fessler, P. Lindner, T. Y. Mathä, C. Rampazzi, F. Savignac, T. Schmidt, M. Schürz and P.Vermeulen, “How do households allocate their assets? Stylised facts from the Eurosystem Household Finance and Consumption Survey,” August 2014

505. Y. Kalantzis, “Financial fragility in small open economies: firm balance sheets and the sectoral structure,” August 2014

506. V. Bignon and R. Dutu, “Coin assaying and commodity money,” August 2014

507. A. Pizzo, “The Shimer puzzle(s) in New Keynesian framework,” September 2014

508. F. Langot, L. Patureau and T. Sopraseuth, “Fiscal Devaluation and Structural Gaps,” September 2014

509. G. Gaballo, “Sequential Coordination, Higher-Order Belief Dynamics and E-Stability Principle,” October 2014

510. G. Bazot, M. D. Bordo and E. Monnet, “The Price of Stability. The balance sheet policy of the Banque de France and the Gold Standard (1880-1914) ,” October 2014

511. K. Istrefi and A. Piloiu, “Economic Policy Uncertainty and Inflation Expectations,” October 2014

512. C. Jardet and A. Monks, “Euro Area monetary policy shocks: impact on financial asset prices during the crisis?,” October 2014

513. J. Dugast and T. Foucault, “False News, Informational Efficiency, and Price Reversals,” October 2014

514. G. Cette, J. Lopez and J. Mairesse, “Product and Labor Market Regulations, Production Prices, Wages and Productivity,” October 2014

515. L. Ferrara and C. Marsilli, “Nowcasting global economic growth: A factor-augmented mixed-frequency approach,” October 2014

516. F. Langot and M. Lemoine, “Strategic fiscal revaluation or devaluation: why does the labor wedge matter? ,” October 2014

517. J.-P. Renne, “Fixed-Income Pricing in a Non-Linear Interest-Rate Model,” October 2014

518. J.-P. Renne, “Options Embedded in ECB Targeted Refinancing Operations,” October 2014

519. A. Penalver, “Pre-crisis credit standards: monetary policy or the savings glut?,” November 2014

520. C. Marsilli, “Variable Selection in Predictive MIDAS Models,” November 2014

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