Bayesian MIDAS Penalized Regressions: Estimation, Selection, and Prediction

Matteo Mogliani[1]

[1]Banque de France - International Macroeconomic Division

January 18th, 2019

The views expressed in this paper are those of the authors and do not necessarily reflect those of the Banque de France

Model selection in MIDAS regressions

- **MIDAS (MIxed-DAta Sampling)** : a regression method that relates a variable measured at some frequency to current and lagged values of variables measured at a *higher* frequency (Ghysels et al., 2004).

- Gained considerable attention in the last decade : proved to improve forecasts of macroeconomic variables, such as GDP (Clements and Galvão, 2008, 2009 ; Armesto et al., 2010 ; Andreou et al., 2013 ; Mogliani et al., 2017).

- Selection of predictors problematic, especially in presence of many high-frequency variables.

- In the literature, a few popular model reduction strategies :
  - U-MIDAS + General-to-specific algorithm : jointly select relevant predictors and relevant high-frequency lags (Castle et al., 2009 ; Castle and Hendry 2010 ; Bec and Mogliani, 2015).
  - Factor-augmented MIDAS : common factors (extracted from high-frequency variables) in MIDAS regressions (Marcellino and Schumacher, 2010).
  - Targeted predictors : pre-selection of relevant high-frequency variables based on hard- and soft-thresholding rules (Bai and Ng, 2008).

Model selection in MIDAS regressions

Recently, more efforts to combine MIDAS regressions and selection techniques based on penalized regressions.

- Marsilli (2014) : functional MIDAS + LASSO objective function, solved in 1-step through non-linear optimisation algorithms (Nesterov, 2005).
- Siliverstovs (2017) : Factor-MIDAS *after* U-MIDAS + Elastic-Net objective function (2-step approach).
- Uematsu and Tanaka (in press) : UMIDAS + LASSO objective function.

Compared to these contributions, we provide MIDAS model specification and parameters estimation in one single, and not too much computationally intensive, step while addressing some issues related to the Lasso in a mixed-frequency framework.

Our proposed approach : simultaneous selection and estimation via penalized regressions

The main ingredients :

- MIDAS framework based on Almon lag polynomials
- Bayesian penalized regressions based on Group Lasso and Spike-and-Slab priors
- Adaptive shrinkage
- Penalty hyper-parameters tuned via approximate EM algorithm

The main results :

- Computationally efficient
- Very reliable variables selection and accurate estimation of model parameters
- Accurate out-of-sample results

Outline

Bayesian MIDAS penalized regressions

Tuning the penalty hyper-parameters

Monte Carlo experiments

Empirical application : forecasting US GDP with high-frequency predictors

MIDAS setup

- y_t low frequency variable (observed at discrete times) ;
- $\mathbf{x}_t^{(m)} = (x_{1,t}^{(m)}, \ldots, x_{K,t}^{(m)})'$ vector of K high frequency variables (observed m times between $t-1$ and t).
- For ease of exposition, I will assume for the moment $\mathbf{K} = \mathbf{1}$, but generalization is straightforward.

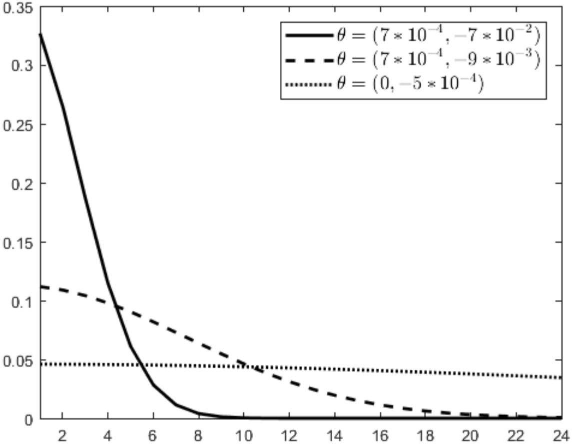
- MIDAS regression :

$$y_t = \alpha + \beta \mathcal{B}\left(L^{1/m}; \boldsymbol{\theta}\right) x_{t-h}^{(m)} + \epsilon_t$$

  - $\mathcal{B}\left(L^{1/m}; \boldsymbol{\theta}\right)$ lag distribution function (sum to 1).
  - $L^{1/m}$ high-frequency lag operator $\Rightarrow L^{1/m} x_t^{(m)} = x_{t-1/m}^{(m)}$.
  - $\boldsymbol{\theta}$ vector of functional parameters.
  - β overall slope coefficient.
  - $h = 0, 1/m, 2/m, 3/m, \ldots$ forecast horizon.

- Several functional forms for $\mathcal{B}\left(L^{1/m}; \boldsymbol{\theta}\right)$ proposed in the literature : exponential Almon or Beta lag polynomials (Ghysels et al., 2007).

Example of MIDAS weights : exponential Almon lag



$$\mathcal{B}\left(L^{1/m};\boldsymbol{\theta}\right) = \sum_{c=1}^{C} B\left(c;\boldsymbol{\theta}\right) L^{c/m} = \sum_{c=1}^{C} \frac{\exp(\theta_1 c + \theta_2 c^2)}{\sum_{c=1}^{C} \exp(\theta_1 c + \theta_2 c^2)} L^{c/m}$$

MIDAS with Almon lag polynomial

- MIDAS with **Almon lag polynomial** $\Rightarrow \mathcal{B}\left(L^{1/m};\boldsymbol{\theta}\right) = \sum_{c=0}^{C-1}\sum_{i=0}^{p}\theta_i c^i L^{c/m}$
- Under the "direct method" for Almon lag polynomials :

$$y_t = \alpha + \boldsymbol{\theta}'\mathbf{Z}_{t-h}^{(m)} + \epsilon_t$$

where $\mathbf{Z}_t^{(m)}$ is a vector of linear combinations of the original high-frequency regressor, $\mathbf{Z}_t^{(m)} = \mathbf{Q}\mathbf{X}_t^{(m)}$, with

- $\mathbf{X}_t^{(m)} = \left(x_t^{(m)}, x_{t-1/m}^{(m)}, \ldots, x_{t-(C-1)/m}^{(m)}\right)'$ a $(C \times 1)$ vector of high frequency data lags
- $\mathbf{Q}$ a $(p+1 \times C)$ polynomial weighting matrix :

$$\mathbf{Q} = \begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ 0 & 1 & 2 & \cdots & (C-1) \\ 0 & 1 & 2^2 & \cdots & (C-1)^2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 1 & 2^p & \cdots & (C-1)^p \end{pmatrix}$$

MIDAS with Almon lag polynomial

Advantages :

- linearity.
- parsimony : $(p+1)$ parameters, *i.e.* lag polynomial coefficients $\boldsymbol{\theta} = (\theta_0, \theta_1, \ldots, \theta_p)$.
- linear zero restrictions : jointly constrain the weighting structure to tail off slowly to zero.
  $\Rightarrow$ obtained by modifying the $\mathbf{Q}$ matrix.
  $\Rightarrow$ the number of parameters reduces from $(p+1)$ to $(p-r+1)$, where $r \leq p$ is the number of restrictions.
- overall slope coefficient computed as $\widehat{\beta} = \boldsymbol{\iota}_C \mathbf{Q}' \hat{\boldsymbol{\theta}}'$, where $\boldsymbol{\iota}_C$ is a $(C \times 1)$ row vector of ones.

Drawback : easily affected by overparameterization and multicollinearity in presence of numerous and potentially correlated predictors.

MIDAS Lasso regression

For **K** $>>$ **1**, we could regularize the model by implementing the selection operator provided by the Lasso (Marsilli, 2014 ; Uematsu and Tanaka, in press).

- Lasso estimator (Tibshirani,1996) with loss function :

$$\mathcal{Q}_{\mathrm{L}}(\boldsymbol{\theta}) = T^{-1}\mathcal{L}_T(\boldsymbol{\theta}) + \lambda\|\boldsymbol{\theta}\|_1$$

  where $\mathcal{L}_T(\boldsymbol{\theta})$ is the negative log-likelihood function and $\|\boldsymbol{\theta}\|_1 = \sum_{k=1}^{K}\sum_{i=0}^{p-r}|\theta_{k,i}|$ denotes the ℓ_1 norm.

- Account for adaptive penalty to achieve the oracle property (Zou, 2006) :

$$\mathcal{Q}_{\mathrm{AL}}(\boldsymbol{\theta}) = T^{-1}\mathcal{L}_T(\boldsymbol{\theta}) + \sum_{k=1}^{K}\sum_{i=0}^{p-r}\lambda_{k,i}|\theta_{k,i}|$$

But Lasso might not be suited in the present framework : lags of high-frequency predictors are by construction highly correlated and hence the Lasso estimator would tend to select randomly only one lag and shrink the remaining polynomial coefficients to zero.

MIDAS Group Lasso regression

We then consider the following penalized regression :

- **Adaptive Group Lasso** (AGL) estimator (Yuan and Lin, 2006 ; Wang and Leng, 2008), where each group is a lag polynomial of size $g_j = p - r + 1$:

$$\mathcal{Q}'_{\mathrm{AGL}}(\boldsymbol{\theta}) = T^{-1}\mathcal{L}_T(\boldsymbol{\theta}) + \sum_{j=1}^{G} \lambda_j \|\boldsymbol{\theta}_j\|_2$$

  where $\|\boldsymbol{\theta}_j\|_2 = (\boldsymbol{\theta}'_j \boldsymbol{\theta}_j)^{1/2}$ denotes the ℓ_2 norm on each partition j of vector $\boldsymbol{\theta}$.

Rationale for the lag polynomial grouping structure : if one high-frequency predictor is irrelevant, it should be expected that zero-coefficients occur in all the parameters of its lag polynomial.

Parsimony : $G = K$ penalty hyper-parameters, instead of $K(p - r + 1)$ as in the Lasso.

Bayesian hierarchical approach for MIDAS Group Lasso (BMIDAS-AGL)

Usual (frequentist) choice for inference relies on Group LARS algorithms (Yuan and Lin, 2006).

Here we consider a **Bayesian hierarchical approach** :

- Exploit model inference via posterior distributions of parameter $\Rightarrow$ use posterior distribution to set a variable selection criterion, *e.g.* :

$$0 \in \text{ credible interval of } \widehat{\beta}_k \Rightarrow \text{irrelevant predictor}$$

- Provide a flexible way of estimating the penalty parameters :

$$\widehat{\boldsymbol{\lambda}} = \underset{\boldsymbol{\lambda}}{\text{argmax}}\ \pi(y|\boldsymbol{\lambda})$$

- Hierarchical priors that lead to conditional posteriors from which we can efficiently draw through MCMC. BMIDAS-AGL

MIDAS Group Lasso with Spike-and-Slab priors (BMIDAS-AGL-SS)

- A sparse solution cannot be actually achieved with the BMIDAS-AGL, as the Bayesian approach provides a shrinkage of the coefficients towards zero, but usually not exactly to zero.
- Add a point mass mixture prior to penalized regressions (Zhang et al., 2014 ; Zhao and Sarkar, 2015 ; Ročková and George, 2018).
- We follow Xu and Ghosh (2015) and we consider **spike-and-slab priors** for variable selection in Group Lasso regressions.
- Point mass at $\mathbf{0}$ (spike part of the prior) and Group Lasso prior on the slab part : facilitates variable selection at the group level and shrinks coefficients in the selected groups.
- Equivalent to adding a ℓ_0-like penalty, which penalizes the number of nonzero groups in the predictors. ▸ BMIDAS-AGL-SS

Outline

Tuning the penalty hyper-parameters

- So far, we have put a prior $\pi(\boldsymbol{\lambda})$, such that :

$$\pi(\phi|y) = \int \pi(\phi|y, \boldsymbol{\lambda})\pi(\boldsymbol{\lambda}|y)d\boldsymbol{\lambda}$$

  where $\phi = (\boldsymbol{\theta}, \boldsymbol{\tau}, \sigma^2)'$. However the posterior distribution $\pi(\boldsymbol{\lambda}|y) = \pi(y|\boldsymbol{\lambda})\pi(\boldsymbol{\lambda})$ may be sensitive to the choice of the prior $\pi(\boldsymbol{\lambda})$.

- A popular alternative approach is represented by the **Monte Carlo EM algorithm**, MCEM (Casella, 2001).
- Treat ϕ as missing data and iterate over N Monte Carlo iterations, where $\boldsymbol{\lambda}$ is **updated** at each nth iteration up to convergence :
  - E-step : for a given $\boldsymbol{\lambda}^{(n)}$, solve

$$Q(\boldsymbol{\lambda}|\boldsymbol{\lambda}^{(n)}) = \int \log\left[f_{\phi,\lambda}(y)\pi(\phi|\boldsymbol{\lambda})\right]\pi(\phi|y, \boldsymbol{\lambda}^{(n)})d\phi$$

  - M-step : maximize $Q(\boldsymbol{\lambda}|\boldsymbol{\lambda}^{(n)})$ to give $\boldsymbol{\lambda}^{(n+1)}$

$$\widehat{\boldsymbol{\lambda}}^{(n+1)} = \underset{\boldsymbol{\lambda}}{\operatorname{argmax}}\ Q(\boldsymbol{\lambda}|\boldsymbol{\lambda}^{(n)})$$

  - $\pi(\phi|y, \boldsymbol{\lambda})$ intractable, hence simulation method required (a run of the Gibbs sampler).

Tuning the penalty hyper-parameters : Approximate EM

- MCEM extremely costly : fully converged Gibbs sampling from $\pi(\phi|y, \boldsymbol{\lambda}^{(n)})$ for each nth Monte Carlo iteration.

- We rely on an **adaptive MCMC algorithm** : stochastic approximation to EM solution (Atchadé, 2011).
- Replacing the full maximization of the Q function by one step of the gradient algorithm :

$$\boldsymbol{\lambda}^{(s+1)} = \boldsymbol{\lambda}^{(s)} + a^{(s)} \nabla_{\boldsymbol{\lambda}} Q(\boldsymbol{\lambda}^{(s)}|\boldsymbol{\lambda}^{(s)})$$

  where $a^{(s)}$ is a step-size.
- Change in superscript : from (n) Monte Carlo iteration to (s) Gibbs sampler iteration $\Rightarrow$ computationally efficient : a single Monte Carlo run $(N = 1)$ is required !

- If we approximate $\nabla_{\boldsymbol{\lambda}} Q(\boldsymbol{\lambda}^{(s)}|\boldsymbol{\lambda}^{(s)})$ with the more tractable $\nabla_{\boldsymbol{\lambda}} \log \pi(\phi^{(s+1)}|\boldsymbol{\lambda}^{(s)})$, we can build upon simple analytic solutions.

Tuning the penalty hyper-parameters : Approximate EM

Updating rules for $\boldsymbol{\lambda}^{(s+1)}$:

- Make the transformation $\boldsymbol{\omega} = \frac{1}{2}\log(\boldsymbol{\lambda})$:

  $$\nabla_{\omega} \log \pi(\phi|\boldsymbol{\omega}) = (\mathbf{g} + 1) - \exp(2\boldsymbol{\omega}) \odot \boldsymbol{\tau}$$

  where $\mathbf{g} = (g_1, \ldots, g_G)'$ and $\odot$ is the element-wise product.

- Updating rule for $\boldsymbol{\omega}$:

  $$\omega_j^{(s+1)} = \omega_j^{(s)} + a^{(s)} \left[(g_j + 1) - \exp\left(2\omega_j^{(s)}\right) \tau_j^{2,(s+1)} \right]$$

Hence $\boldsymbol{\lambda}^{(s+1)} = \exp(2\boldsymbol{\omega}^{(s+1)})$.

Numerical illustration

- MIDAS DGP with exponential Almon function and four predictors $(K = 4)$:

$$y_t = \beta_0 + \sum_{k=1}^{4} \beta_k \sum_{c=1}^{C} B\left(c; \boldsymbol{\vartheta}\right) L^{c/m} x_{k,t}^{(m)} + \epsilon_t$$

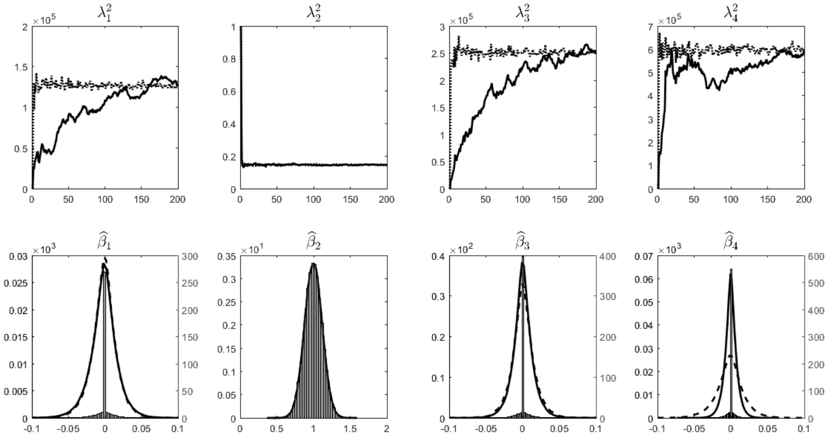
$$B\left(c; \boldsymbol{\vartheta}\right) = \frac{\exp(\vartheta_1 c + \vartheta_2 c^2)}{\sum_{c=1}^{C} \exp(\vartheta_1 c + \vartheta_2 c^2)}$$

  with $(\beta_1, \beta_2, \beta_3, \beta_4) = (0, 1, 0, 0)$, $C = 12$, $m = 3$, $\boldsymbol{\vartheta} = (0.10, -0.15) \Rightarrow$ fast-decaying weights.

- We estimate our models with $p = 3$ and $r = 2$.

- We tune $\boldsymbol{\lambda}$ using both methods :
  - Stochastic approximation with $N = 1$ and $S = 400,000$
  - MCEM with $N = 200$ and $S = 50,000$

Numerical illustration



The first panel illustrates the evolution of the penalty hyper-parameters $\boldsymbol{\lambda}$ across iterations of the stochastic approximation approach for BMIDAS-AGL (solid lines) and BMIDAS-AGL-SS model (dotted lines). The second panel illustrates the posterior distributions of parameters $\boldsymbol{\beta}$ for BMIDAS-AGL (solid lines) and BMIDAS-AGL-SS model (histogram) using the stochastic approximation approach, and the BIMDAS-AGL model using the MCEM algorithm (dashed lines).

- Computation time : 2 minutes SA *vs* 30 minutes MCEM !

Outline

Design of the experiments

$$y_t = \alpha + \sum_{k=1}^{K} \beta_k \sum_{c=1}^{C} B(c; \boldsymbol{\vartheta}) L^{c/3} x_{k,t-h}^{(3)} + \epsilon_t$$

$$x_{k,t}^{(3)} = \mu + \rho x_{k,t-1/3}^{(3)} + \varepsilon_{k,t} \qquad B(c; \boldsymbol{\vartheta}) = \frac{\exp(\vartheta_1 c + \vartheta_2 c^2)}{\sum_{c=1}^{C} \exp(\vartheta_1 c + \vartheta_2 c^2)}$$

- $T = 200$
- $K = \{30, 50\}$
- $\boldsymbol{\beta} = (0, 0.3, 0.5, 0, 0.3, 0.5, 0, 0, 0.8, \mathbf{0})'$
- $\boldsymbol{\vartheta}_1 = (7 * 10^{-4}, -7 * 10^{-2}) \Rightarrow$ fast-decaying weights
  $\boldsymbol{\vartheta}_2 = (7 * 10^{-4}, -9 * 10^{-3}) \Rightarrow$ slow-decaying weights
  $\boldsymbol{\vartheta}_3 = (0, -5 * 10^{-4}) \Rightarrow$ near-flat weights
- ϵ_t and $\boldsymbol{\varepsilon}_t$ i.i.d. with distribution :

$$\begin{pmatrix} \epsilon_t \\ \boldsymbol{\varepsilon}_t \end{pmatrix} \sim \text{i.i.d.}\mathcal{N}\left[\begin{pmatrix} 0 \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \sigma^2 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_\varepsilon \end{pmatrix}\right],$$

where $\boldsymbol{\Sigma}_\varepsilon$ has elements $\sigma_\varepsilon^{|k-k'|}$, with $\sigma_\varepsilon = \{0.50, 0.95\}$, and σ fixed such that SNR $= 0.20$.

Simulation results : variable selection performance

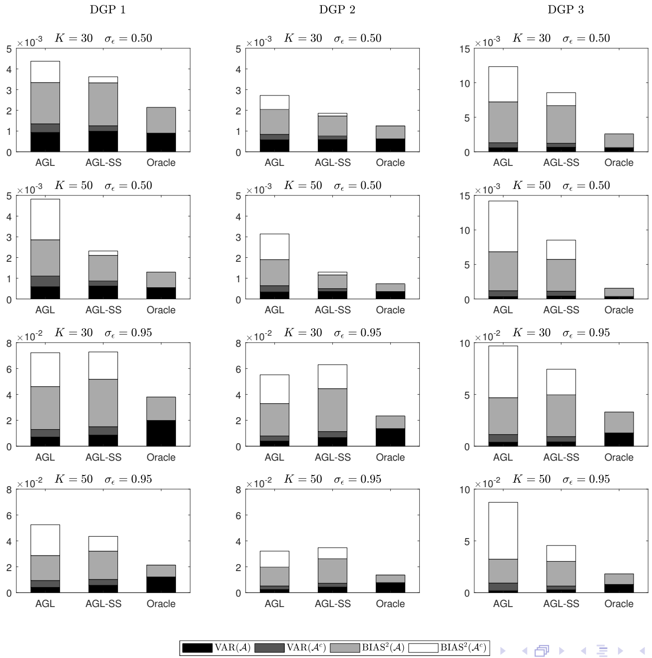
K	σ_ε	TPR	FPR	MCC	TPR	FPR	MCC	TPR	FPR	MCC
		BMIDAS-AGL								
		DGP 1			DGP 2			DGP 3		
30	0.50	**0.95**	0.03	0.90	0.94	0.01	0.94	0.85	0.14	0.63
	0.95	**0.37**	0.04	0.42	**0.46**	0.05	0.48	**0.36**	0.11	0.29
50	0.50	0.92	0.04	0.82	0.96	0.03	0.86	0.77	0.16	0.46
	0.95	**0.33**	0.04	0.36	**0.45**	0.03	0.49	0.26	0.10	0.16
		BMIDAS-AGL-SS								
		DGP 1			DGP 2			DGP 3		
30	0.50	0.94	**0.01**	**0.94**	**0.98**	**0.01**	**0.96**	**0.86**	**0.07**	**0.76**
	0.95	0.33	**0.03**	**0.42**	0.41	**0.03**	**0.48**	0.32	**0.05**	**0.36**
50	0.50	**0.94**	**0.01**	**0.94**	**0.98**	**0.01**	**0.97**	**0.80**	**0.06**	**0.69**
	0.95	0.31	**0.01**	**0.44**	0.43	**0.01**	**0.55**	**0.29**	**0.03**	**0.37**

DGP 1 : $\vartheta_1 = (7 * 10^{-4}, -7 * 10^{-2})$

DGP 2 : $\vartheta_2 = (7 * 10^{-4}, -9 * 10^{-3})$

DGP 3 : $\vartheta_3 = (0, -5 * 10^{-4})$

Simulation results : breakdown of MSE by active ($\mathcal{A}$) and inactive set ($\mathcal{A}^c$)



Simulation results : out-of-sample performance

K	σ_ε	MSFE	-LS	CRPS	MSFE	-LS	CRPS	MSFE	-LS	CRPS
		BMIDAS-AGL								
		DGP 1			DGP 2			DGP 3		
30	0.50	2.07	1.76	0.81	1.52	1.61	0.69	**1.56**	**1.63**	**0.71**
	0.95	**4.80**	2.20	1.24	3.75	2.09	1.10	4.26	2.16	1.18
50	0.50	2.25	1.85	0.86	1.73	1.80	0.73	1.77	1.71	0.76
	0.95	**6.32**	**2.35**	**1.43**	**3.24**	**2.00**	**1.01**	**4.68**	**2.19**	**1.21**
		BMIDAS-AGL-SS								
		DGP 1			DGP 2			DGP 3		
30	0.50	**1.84**	**1.71**	**0.77**	**1.47**	**1.60**	**0.68**	1.62	1.65	0.72
	0.95	4.90	**2.19**	**1.24**	**3.75**	**2.09**	**1.10**	**4.23**	**2.14**	**1.16**
50	0.50	**2.18**	**1.83**	**0.84**	**1.56**	**1.65**	**0.71**	**1.54**	**1.63**	**0.71**
	0.95	6.54	2.37	1.44	3.50	2.03	1.05	4.73	2.19	1.22

DGP 1 : $\vartheta_1 = (7 * 10^{-4}, -7 * 10^{-2})$

DGP 2 : $\vartheta_2 = (7 * 10^{-4}, -9 * 10^{-3})$

DGP 3 : $\vartheta_3 = (0, -5 * 10^{-4})$

Simulation results : difference in variable selection performance $r = 0$ vs $r = 2$

K	σ_ε	TPR	FPR	MCC	TPR	FPR	MCC	TPR	FPR	MCC
		BMIDAS-AGL								
		DGP 1			DGP 2			DGP 3		
30	0.50	0.16	-0.02	0.16	0.09	-0.01	0.08	-0.08	0.10	-0.24
	0.95	0.06	-0.01	**0.08**	0.10	0.00	0.08	-0.06	0.06	-0.18
50	0.50	0.22	-0.03	0.21	0.17	-0.03	0.19	**-0.07**	0.10	-0.26
	0.95	0.09	0.00	0.09	0.11	0.00	0.12	-0.13	0.07	-0.26
		BMIDAS-AGL-SS								
		DGP 1			DGP 2			DGP 3		
30	0.50	**0.12**	**-0.01**	**0.11**	**0.06**	**0.00**	**0.05**	**-0.08**	**0.05**	**-0.17**
	0.95	**0.05**	**-0.01**	0.09	**0.05**	**0.00**	**0.04**	**-0.05**	**0.01**	**-0.07**
50	0.50	**0.05**	**0.00**	**0.02**	**0.08**	**0.00**	**0.07**	-0.13	**0.05**	**-0.23**
	0.95	**0.04**	**0.00**	**0.05**	**0.08**	**0.00**	**0.09**	**-0.09**	**0.01**	**-0.13**

$$\text{DGP 1} : \boldsymbol{\vartheta}_1 = (7 * 10^{-4}, -7 * 10^{-2})$$

$$\text{DGP 2} : \boldsymbol{\vartheta}_2 = (7 * 10^{-4}, -9 * 10^{-3})$$

$$\text{DGP 3} : \boldsymbol{\vartheta}_3 = (0, -5 * 10^{-4})$$

Outline

Bayesian MIDAS penalized regressions

Tuning the penalty hyper-parameters

Monte Carlo experiments

Empirical application : forecasting US GDP with high-frequency predictors

The data

- Annualized q-o-q growth rate of GDP (1980Q1-2017Q4)
  - Out-of-sample : 2000Q1-2017Q4
- Monthly predictors (33) :
  - macroeconomic series extracted from the FRED-MD database (output and income, labor, housing, consumption, and orders).
- Weekly predictors (3) :
  - Chicago Fed financial conditions indicator (risk, leverage, and credit subcomponents)
- Daily predictors (6) :
  - effective Federal Funds rate
  - 10-year government bond rate − FF rate
  - ADS daily business cycle indicator
  - returns on the portfolio of small minus big stocks
  - returns on the portfolio of high minus low book-to-market ratio stocks
  - returns on a winner minus loser momentum spread portfolio

$T = 152$, $T_{oos} = 72$, and $K = 42$.

Benchmark and alternative models

- Random-walk (benchmark)
- AR(1) model.
- Combination of K single-indicator Bayesian MIDAS (BMIDAS-comb) as in Pettenuzzo et al. (2016), where the combination weights are computed using a discounted version of the optimal prediction pool proposed by Geweke and Amisano (2011).
- Bayesian model selection (BMS) as in Lamnisos et al. (2013), where an Adaptive Metropolis-Hastings algorithm is implemented to tune automatically the model proposals and achieve a targeted acceptance rate. We select those variables displaying a posterior probability of inclusions greater than 50%.
- Bayesian model averaging (BMA), estimated through a reversible-jump MC^3 algorithm with g-BRIC prior.

Note that all the models (including BMIDAS-AGL and BMIDAS-AGL-SS) include one lag of the dependent variable and the same Almon lag polynomial structure. BMS and BMA are modified to account for lag polynomial group selection.

Out-of-sample results

Out-of-sample : 2000Q1-2017Q4

	$h = 0$			$h = 1$			$h = 4$		
	ΔRMSFE	ΔLS	ΔCRPS	ΔRMSFE	ΔLS	ΔCRPS	ΔRMSFE	ΔLS	ΔCRPS
BMIDAS-AGL	0.61	0.54	0.59	0.74	0.33	0.72	0.82	0.24	0.81
	(**0.00**)	(**0.00**)	(**0.00**)	(**0.00**)	(**0.00**)	(**0.00**)	(**0.10**)	(**0.07**)	(**0.06**)
BMIDAS-AGL-SS	**0.57**	**0.58**	**0.56**	**0.70**	**0.39**	**0.68**	0.81	0.24	0.77
	(**0.00**)	(**0.00**)	(**0.00**)	(**0.00**)	(**0.00**)	(**0.00**)	(**0.10**)	(**0.06**)	(**0.03**)
AR(1)	0.85	0.16	0.82	0.85	0.16	0.82	0.80	0.22	**0.76**
BMIDAS-comb	0.66	0.44	0.65	0.77	0.30	0.74	**0.78**	**0.27**	0.76
BMS	1.18	-0.67	1.32	1.12	-0.15	1.15	0.85	0.18	0.87
BMA	0.61	0.48	0.59	0.76	0.30	0.73	0.84	0.16	0.80

Notes : predictive performance of model i compared to the random-walk benchmark. Bold values denote the best outcomes. In parentheses, p-values for the test of the null hypothesis of equal predictive accuracy at 10% level according to the one-sided t-statistic version of the DMW test.

Concluding remarks

- A new approach to select and estimate simultaneously mixed-frequency regressions.
- BMIDAS-AGL and BMIDAS-AGL-SS models provide very good simulation results in terms of selection accuracy. However, an improvement when dealing with highly correlated predictors is needed.
- Approximate EM algorithm crucial to obtain computational efficiency.
- When applied to US GDP, all in all our models display good nowcasting performance.
- Potential (future) extensions to time-varying MIDAS regressions and quantile MIDAS regressions.

Bayesian hierarchical model : BMIDAS-AGL

- Following Kyung et al. (2010), the conditional prior for $\boldsymbol{\theta}$ is :

$$\pi(\boldsymbol{\theta}|\sigma^2) \propto \exp\left(-\frac{1}{\sqrt{\sigma^2}}\sum_{j=1}^{G}\lambda_j\|\boldsymbol{\theta}_j\|_2\right)$$

- Hierarchical prior :

$$y|\mathbf{Z},\boldsymbol{\theta},\sigma^2 \sim \mathcal{N}\left(\boldsymbol{\theta}'\mathbf{Z},\sigma^2\mathbf{I}_T\right)$$

$$\boldsymbol{\theta}_j|\tau_j^2,\sigma^2 \sim \mathcal{N}(\mathbf{0},\sigma^2\tau_j^2\mathbf{I}_{g_j}) \qquad j=1,\ldots,G$$

$$\tau_j^2 \sim \mathrm{Gamma}\left(\frac{g_j+1}{2},\frac{\lambda_j^2}{2}\right)$$

$$\sigma^2 \sim \mathrm{iGamma}\left(a_1,b_1\right)$$

where $\boldsymbol{\tau}=(\tau_1^2,\ldots,\tau_G^2)$, $\boldsymbol{\lambda}=(\lambda_1^2,\ldots,\lambda_G^2)$, and $\mathbf{I}_{g_j}$ is the identity matrix of order g_j.

Bayesian hierarchical model : BMIDAS-AGL

- Full posterior distribution of all the unknown parameters conditional on the data and the penalty hyper-parameters :

$$\pi(\boldsymbol{\theta}, \boldsymbol{\tau}, \sigma^2 | \boldsymbol{\lambda}, y, \mathbf{Z}) \propto \left(\sigma^2\right)^{-\frac{T+\tilde{g}-1}{2}-a_1-1} \exp\left[-\frac{1}{2\sigma^2}\|y-\boldsymbol{\theta}'\mathbf{Z}\|_2^2 - \frac{b_1}{\sigma^2}\right]$$

$$\times \prod_{j=1}^{G}\left(\frac{1}{2\pi\sigma^2\tau_j^2}\right)^{\frac{g_j}{2}} \exp\left(-\frac{\|\boldsymbol{\theta}_j\|_2^2}{2\sigma^2\tau_j^2}\right)$$

$$\times \prod_{j=1}^{G}\left(\lambda_j^2\right)^{\frac{g_j+1}{2}} \left(\tau_j^2\right)^{\frac{g_j+1}{2}-1} \exp\left(-\frac{\lambda_j^2}{2}\tau_j^2\right)$$

Bayesian hierarchical model : BMIDAS-AGL

- Full conditional posteriors, with $\lambda_j^2 \sim \mathrm{Gamma}\left(a_2, b_2\right)$:

$$\boldsymbol{\theta}_j | \text{rest} \sim \mathcal{N}\left(\mathbf{A}_j^{-1}\mathbf{C}_j, \sigma^2\mathbf{A}_j^{-1}\right)$$

$$\tau_j^{-2} | \text{rest} \sim \mathrm{iGaussian}\left(\frac{\lambda_j \sigma}{\|\boldsymbol{\theta}_j\|_2}, \lambda_j^2\right)$$

$$\sigma^2 | \text{rest} \sim \mathrm{iGamma}\left(\frac{T + \tilde{g} - 1}{2} + a_1, \frac{1}{2}\|y - \boldsymbol{\theta}'\mathbf{Z}\|_2^2 + \frac{1}{2}\sum_{j=1}^{G}\frac{\|\boldsymbol{\theta}_j\|_2^2}{\tau_j^2} + b_1\right)$$

$$\lambda_j^2 | \text{rest} \sim \mathrm{Gamma}\left(\frac{g_j + 1}{2} + a_2, \frac{\tau_j^2}{2} + b_2\right)$$

where $\mathbf{A}_j = \mathbf{Z}_j'\mathbf{Z}_j + \tau_j^{-2}\mathbf{I}_{g_j}$, $\mathbf{C}_j = \mathbf{Z}_j'\left(y - \boldsymbol{\theta}_{-j}'\mathbf{Z}_{-j}\right)$, and
$\boldsymbol{\theta}_{-j} = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_{j-1}, \boldsymbol{\theta}_{j+1}, \ldots, \boldsymbol{\theta}_G)'$.

return

Bayesian hierarchical model : BMIDAS-AGL-SS

- Hierarchical prior

$$y|\mathbf{Z},\boldsymbol{\theta},\sigma^2 \sim \mathcal{N}\left(\boldsymbol{\theta}'\mathbf{Z},\sigma^2\mathbf{I}_T\right)$$

$$\boldsymbol{\theta}_j|\tau_j^2,\sigma^2,\pi_0 \sim (1-\pi_0)\mathcal{N}(\mathbf{0},\sigma^2\tau_j^2\mathbf{I}_{g_j}) + \pi_0\delta_0(\boldsymbol{\theta}_j) \qquad j=1,\ldots,G$$

$$\tau_j^2 \sim \mathrm{Gamma}\left(\frac{g_j+1}{2},\frac{\lambda_j^2}{2}\right)$$

$$\sigma^2 \sim \mathrm{iGamma}\left(a_1,b_1\right)$$

$$\pi_0 \sim \mathrm{Beta}\left(c,d\right)$$

where $\delta_0(\boldsymbol{\theta}_j)$ denotes a point mass at $\mathbf{0} \in \mathbb{R}^{g_j}$. We follow Castillo et al. (2015) and Ročková and George (2018), and we set $c = 1$ and $d = G$.

Bayesian hierarchical model : BMIDAS-AGL-SS

- Full posterior distribution of all the unknown parameters conditional on the data and the penalty hyper-parameters :

$$\pi(\boldsymbol{\theta}, \boldsymbol{\tau}, \sigma^2, \pi_0 | \boldsymbol{\lambda}, y, \mathbf{Z}) \propto (\sigma^2)^{-\frac{T+\tilde{g}-1}{2}-a_1-1} \exp\left[-\frac{1}{2\sigma^2}\|y-\boldsymbol{\theta}'\mathbf{Z}\|_2^2 - \frac{b_1}{\sigma^2}\right]$$

$$\times \prod_{j=1}^{G}\left[\pi_0\left(\frac{1}{2\pi\sigma^2\tau_j^2}\right)^{\frac{g_j}{2}} \exp\left(-\frac{\|\boldsymbol{\theta}_j\|_2^2}{2\sigma^2\tau_j^2}\right)\mathbf{I}_{\boldsymbol{\theta}_j\neq 0}\right.$$

$$\left. + (1-\pi_0)\delta_0(\boldsymbol{\theta}_j)\right]$$

$$\times \prod_{j=1}^{G}(\lambda_j^2)^{\frac{g_j+1}{2}}(\tau_j^2)^{\frac{g_j+1}{2}-1}\exp\left(-\frac{\lambda_j^2}{2}\tau_j^2\right)$$

$$\times \pi_0^{c-1}(1-\pi_0)^{d-1}$$

Bayesian hierarchical model : BMIDAS-AGL-SS

- Full conditional posteriors with $\lambda_j^2 \sim \text{Gamma}(a_2, b_2)$:

$$\boldsymbol{\theta}_j|\text{rest} \sim \gamma_j \, \mathcal{N}\left(\mathbf{A}_j^{-1}\mathbf{C}_j, \sigma^2\mathbf{A}_j^{-1}\right) + (1-\gamma_j)\, \delta_0(\boldsymbol{\theta}_j)$$

$$\tau_j^{-2}|\text{rest} \sim \gamma_j \, \text{iGaussian}\left(\frac{\lambda_j\sigma}{\|\boldsymbol{\theta}_j\|_2}, \lambda_j^2\right) + (1-\gamma_j)\, \text{Gamma}\left(\frac{g_j+1}{2}, \frac{\lambda_j^2}{2}\right)$$

$$\sigma^2|\text{rest} \sim \text{iGamma}\left(\frac{T+\widetilde{G}-1}{2} + a_1, \frac{1}{2}\|y - \boldsymbol{\theta}'\mathbf{Z}\|_2^2 + \frac{1}{2}\sum_{j=1}^{G}\frac{\|\boldsymbol{\theta}_j\|_2^2}{\tau_j^2} + b_1\right)$$

$$\gamma_j|\text{rest} \sim \text{Bernoulli}(\pi_j)$$

$$\lambda_j^2|\text{rest} \sim \text{Gamma}\left(\frac{g_j+1}{2} + a_2, \frac{\tau_j^2}{2} + b_2\right)$$

$$\pi_0|\text{rest} \sim \text{Beta}\left(\sum_{j=1}^{G}\gamma_j + c, \sum_{j=1}^{G}(1-\gamma_j) + d\right)$$

where

$$\pi_j = \pi(\boldsymbol{\theta}_j \neq \mathbf{0}|\text{rest}) = \frac{\pi_0\left[(\tau_j^2)^{-\frac{g_j}{2}}|\mathbf{A}_j|^{-\frac{1}{2}}\exp\left(\frac{1}{2\sigma^2}\mathbf{C}_j'\mathbf{A}_j^{-1}\mathbf{C}_j\right)\right]}{1-\pi_0\left[1-(\tau_j^2)^{-\frac{g_j}{2}}|\mathbf{A}_j|^{-\frac{1}{2}}\exp\left(\frac{1}{2\sigma^2}\mathbf{C}_j'\mathbf{A}_j^{-1}\mathbf{C}_j\right)\right]}.$$