Forecasting with mixed frequency data based on a Bayesian nested lasso regression model

George Michailidis

Department of Statistics and the Informatics Institute
University of Florida

Joint work with
Satyajit Ghosh and Kshitij Khare

Motivating Application: Forecasting GDP - I

- Gross Domestic Product forecasts available at quarterly frequency
- For the US, the first quarterly estimate becomes available four weeks after the end of the corresponding quarter
  For the Euro zone the first quarterly estimate comes at around six weeks
- A number of leading and coincident economic indicators are available at a monthly or higher frequency with no or little publishing delays
- Beneficial to diverse stakeholders to build forecasting models that leverage such data and also provide near term (one or two months ahead) forecasts

Motivating Application: Forecasting GDP - II

- The GDPNow project by the Atlanta Fed aims to provide frequent updates for the US GDP
- The GDPNow methodology segments the GDP contributors to 13 multivariate components (e.g., personal exenditures for goods and services, federal and state consumptions and investments, etc.) and builds a Bayesian Vector Autoregressive (BVAR) model for each of them
- The BVAR produces quarterly forecasts that are combined with monthly time series autoreregressive based forecasts through regression models
- Forecasting problems with mixed frequency data motivated work in Mixed Data Sampling (MIDAS) regression models (more details in the sequel)

Motivating Application: Forecasting GDP - III

A sneak peak at produced forecasts by the proposed Bayesian methodology

Forecasts of U.S. GDP using lagged values of 125 key economic indicators (details in the sequel)

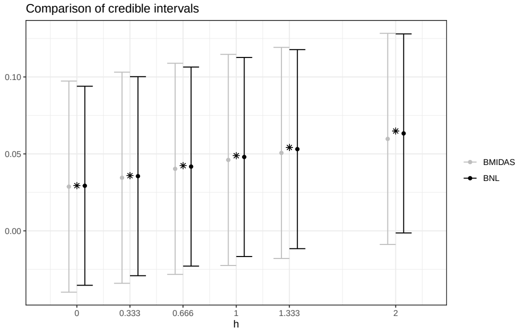


Figure: Credible intervals based on Bayesian Nested Lasso (BNL) and Bayesian MIDAS (BMIDAS) for h-step ahead annualized GDP growth rate forecasts (from $h = 0$ to $h = 2$ quarters) using data up to 2017 Q2. The official BEA values are provided as black asterisks (mid-quarter values are obtained by linear interpolation). The point estimates (posterior median values) corresponding to BNL and BMIDAS are plotted as black and grey dots, respectively.

Motivating Application: Forecasting GDP - IV

Note that the GDP estimate for Q3 2017 (after all revisions have been made that can occur up to three years after the initial estimate, came in at 4.9% ($h = 1$ in the plot), fairly close to the mean prediction of 4.8%

However, the original estimates by the Bureau of Economic Analysis of the US Department of Commerce that is officially responsible to produce such estimates came in at

1. 5.40%; 1 month after the end of Q2
2. 5.65%; 2 months after the end of Q2
3. 5.44%; 3 months after the end of Q2

Mixed Frequency Data Regression Model - I

- Suppose the response variable $\{y_t\}$ is sampled at a low frequency (quarterly), whereas the K predictor variables $\{x_t^i\}_{i=1}^K$ are sampled at a higher frequency (monthly, weekly etc.).
- In particular, for every $1 \leq i \leq K$, there are m_i observations of x^i for a single observation of the response y.
- We want to "project y_t onto a history of lagged observations of itself, as well as $\{x_t^i\}_{i=1}^K$"; i.e.

$$y_t = \sum_{i=1}^{d} \alpha_i y_{t-i} + \sum_{i=1}^{K} \sum_{j=0}^{p_i} \beta_j^i x_{t-j/m_i}^i + \varepsilon_t. \quad (1)$$

High frequency pattern for a quarterly response y, and a monthly predictor x^i

$$\begin{array}{llll} y_{t-d} \quad \cdots \quad y_{t-1} & & & <-- \text{Low frequency AR regressors} \\ \quad \cdots \quad \cdots & \underbrace{(x_{t-2/3}^i, x_{t-1/3}^i, x_t^i)}_{\text{3 per quarter}} & & <-\text{Higher frequency regressors} \end{array}$$

An illustrative example

Consider a response variable y observed at a quarterly frequency.
Further, consider a predictor x observed at a monthly frequency.

The model postulates that for quarter t, we model y_t as a linear combination of the monthly values of x within that quarter; namely $(x_{3t}, x_{3(t-1)}, x_{3(t-2)}$, plus its lagged values.

Hence, we can write the following linear regression representation

$$\begin{bmatrix} y_2 \\ y_3 \\ \dots \\ y_n \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_{n-1} \end{bmatrix} \alpha_1 + \begin{bmatrix} x_3 & x_2 & x_1 \\ x_6 & x_5 & x_4 \\ \dots & \dots & \dots \\ x_{3n} & x_{3n-1} & x_{3n-2} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} + \begin{bmatrix} 1 \\ 2 \\ \dots \\ n \end{bmatrix}$$

The β coefficients can be estimated by least squares.
However, with a larger number of lags and including more variables, there is a proliferation of parameters to be estimated that may exceed the number of available samples.

Brief overview of the MIDAS Literature

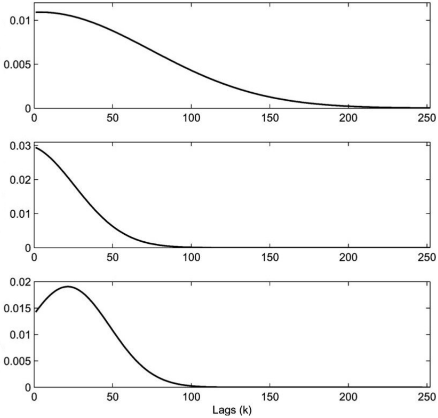
- Ghysels et al. (2007) proposed the use of distributed lag polynomials that involve few parameters and can be estimated using nonlinear least squares algorithms.
- Functional lag polynomials are chosen for β_j^i's to avoid parameter proliferation for long high-frequency lags p_i.
- A popular functional form of the polynomial is the exponential Almon lag and Beta Almon lag given by

$$\begin{array}{lll} \text{MIDAS with Exp. Almon Lag} & \beta_j^i & \dfrac{\exp\left(\eta_1^i j + \eta_2^i j^2\right)}{\sum_{k=0}^{p_i} \exp\left(\eta_1^i k + \eta_2^i k^2\right)} \\ \text{MIDAS with Beta Almon Lag} & \beta_j^i & \dfrac{j^{\eta_1^i - 1}(p_i + 1 - j)^{\eta_2^i - 1}}{\sum_{k=0}^{p_i} k^{\eta_1^i - 1}(p_i + 1 - k)^{\eta_2^i - 1}}, \end{array}$$

with parameters $\eta^i = (\eta_1^i, \eta_2^i)$, respectively.

Illustration of Almon Lag Models

For different values of η_1 and η_2 we get different behavior of the regression coefficients



Mixed Frequency Data Regression Model - II

A key insight provided by the MIDAS methodology is that for mixed frequency regression, a small number of adjacent lags of the predictors can lead to good prediction.

Hence, the following aspects are desirable to be incorporated in the modeling framework:

1. Select a small number of coefficients for adjacent lags of the high-frequency predictor as significant, and set the rest to zero.
2. The regression *coe cients* of further in the past lags should *decay*; i.e., $\beta_j \downarrow$ as $j \uparrow$.

Objective

Develop an efficient Bayesian methodology to forecast a low frequency variable based on its own lag and available high frequency variables.

We also address the following issues

- proper lag selection
- reduce the # of parameter effectively
- maintain the β decay

e.g. forecast **quarterly** observed 'GDP' $\{y_t\}$ based on linear combination of **monthly** observed 'employment payroll' $\{x_{t:1}, x_{t:2}, x_{t:3}\}$ i.e. at the t-th quarter

$$y_t = \begin{pmatrix} x_{t:1} & x_{t:2} & x_{t:3} & x_{t-1:1} \cdots \end{pmatrix}_{p \times 1} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{pmatrix}_{p \times 1} + \varepsilon_t.$$

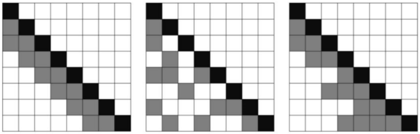
- β_i decays as i (i.e. the lag) increases.
- After lag d, the effect of $x_{t:.}$ becomes insignificant and hence all the $\beta_j = 0 \quad j > d$.

Background on the Nested Lasso

- Levina et al. (2008) proposed the "nested lasso" penalty in the context of large covariance estimation.
- This is useful when a natural ordering of the variables is available.
- $\boldsymbol{X} = (X_1, \cdots, X_p) \quad N_p(\mathbf{0}, \Sigma)$ with $\Sigma^{-1} = T \; DT$.
- T is the unique unit lower triangular matrix with 1 on its diagonal and (i, j)-th element - $\varphi_{i,j}$ for $j < i$, and $D = \text{diag}\,(\sigma_i^2)$.
- Minimize $(\Sigma, \boldsymbol{x}_1, \cdots, \boldsymbol{x}_n) + \sum_2^p Q(j)$ where

$$Q(j) := \lambda \left(|\varphi_{j,j-1}| + \frac{|\varphi_{j,j-2}|}{|\varphi_{j,j-1}|} + \frac{|\varphi_{j,j-3}|}{|\varphi_{j,j-3}|} + \cdots + \frac{|\varphi_{j,1}|}{|\varphi_{j,2}|} \right) \qquad \text{where } 0/0 := 0.$$

E. LEVINA, A. ROTHMAN AND J. ZHU



(a) (b) (c)

The placement of zeros in the Cholesky factor T: (a) Banding; (b) Lasso penalty of Huang et al.; (c) Adaptive banding.

Key insight from the Nested Lasso

To introduce sparsity in $\boldsymbol{\beta} \quad \mathbb{R}^p$ such that there is a "sparsity cutoff" (entries up to a certain index are non-zero, and all entries following that index are zero), then minimizing an objective function with the penalty $|\beta_1| + \sum_{i=2}^{p} |\beta_i / \beta_{i-1}|$ will achieve the goal.

Bayesian Nested Lasso - I

Rewrite the mixed frequency model as

$$\boldsymbol{y}_{n\times 1} = X_{n\times p}\Psi + \boldsymbol{\varepsilon} \qquad \text{with } \boldsymbol{\varepsilon} \quad N_n(\mathbf{0}, \sigma_\varepsilon^2 \mathbf{I}_n). \tag{2}$$

Here n is the number of time points for which the data is observed, $p = d + \sum_{i=1}^{K}(p_i + 1)$, the design matrix X has t-th row

$$(y_{t-1}, \cdots, y_{t-d}, x_t^1, x_{t-1/m_1}^1, \cdots, x_{t-p_1/m_1}^1, \cdots, x_t^K, x_{t-1/m_K}^K, \cdots, x_{t-p_K/m_K}^K),$$

$\alpha = (\alpha_j)_{j=1}^d$, $\boldsymbol{\beta}^i = (\beta_j^i)_{j=0}^{p_i}$ is the vector of lag coefficients corresponding to the i^{th} predictor for $1 \le i \le K$, and $\Psi = (\alpha, \boldsymbol{\beta}^1, \boldsymbol{\beta}^2, \cdots, \boldsymbol{\beta}^K)$ is the combined vector of all regression coefficients. Since we want to introduce a sparsity cutoff in each $\boldsymbol{\beta}^i$, we re-parameterize $\boldsymbol{\beta}^i$ as follows for every $1 \le i \le K$:

$$\begin{pmatrix} \beta_0^i \\ \beta_1^i \\ \beta_2^i \\ \vdots \\ \beta_{p_i}^i \end{pmatrix} = \begin{pmatrix} \theta_0^i \\ \theta_0^i\theta_1^i \\ \theta_0^i\theta_1^i\theta_2^i \\ \vdots \\ \prod_{j=0}^{p_i} \theta_j^i \end{pmatrix}_{p\times 1}. \tag{3}$$

Bayesian Nested Lasso - II

Note that if

$$\theta_k^i = \beta_k^i / \beta_{k-1}^i = 0$$

with $\beta_{-1} = 1$, then all the β_j^i become 0 for all $j \geq k$.

Hence, unstructured sparsity in the θs will induce a sparsity cutoff in the βs and thereby facilitate lag selection.

Specification of Prior Distribution for BNL

- Introduce latent variable $y_i = 1(\theta_i = 0)$.
- We use the following spike-and-slab prior - $BNL(\tau_1, \tau_0, q, a, b)$

$$
\text{For each } 1 \le i \le p \quad
\begin{aligned}
&\theta_i | (\sigma_\varepsilon^2, y_i = 1) \quad N\left(0, \sigma_\varepsilon^2 \tau_1^2\right) \\
&\theta_i | (\sigma_\varepsilon^2, y_i = 0) \quad N\left(0, \sigma_\varepsilon^2 \tau_0^2\right) \\
&P(y_i = 1) = 1 - P(y_i = 0) = q \\
&\sigma_\varepsilon^2 \quad \mathrm{IG}(a, b/2).
\end{aligned}
$$

- The error variance parameter σ_ε^2 is given an $\mathrm{IG}(a, b)$ prior (Inverse-Gamma with shape and rate parameter a and $b/2$ respectively)
- The autoregressive coefficients α_l are given a $N\left(0, \sigma_\varepsilon^2 \tau_\alpha^2\right)$ prior distribution

Modeling Decay in the lag coefficients through the slab variance

The BNL prior is applicable to any regression problem, wherein the variables exhibit a natural ordering (e.g., lags)

However, in the mixed frequency setting, for many variables it is expected that the magnitude of the regression coefficients decreases as a function of the lag

To that end, we specify the slab variance of θ_j^i (now depending on j) through the following generalized logistic function:

$$\tau_1^2(j) := (1 + c)\, \tau_1^2(0)\, \frac{e^{-Bj}}{c + e^{-Bj}},$$

wherein $\tau_1^2(0)$ is the initial slab variance, and the parameters B and c control the rate and shape of decay, respectively.

This specification of the slab variances has only 3 ($\tau_1(0), B$ and c) additional hyper-parameters. The above logistic functional form is flexible enough, since by appropriately selecting values for c and B, slab variance $\tau_1^2(j)$ can decay slowly or fairly fast with respect to lag j.

Numerical illustration of the generalized logistic function

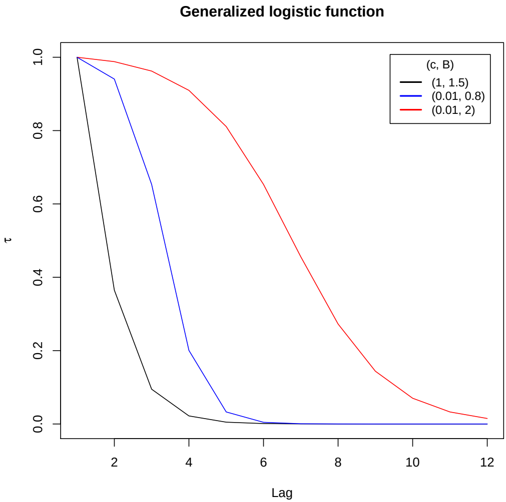


Figure: $\tau_1^2(j)$ with 5 choices of c and B

Final form of the BNL Prior Distribution - II

$$\theta_j^i | (\sigma_\varepsilon^2, \gamma_j^i = 1) \quad N(0, \sigma_\varepsilon^2 \tau_1^2(j))$$

$$\theta_j^i | (\sigma_\varepsilon^2, \gamma_j^i = 0) \quad N(0, \sigma_\varepsilon^2 \tau_0^2)$$

$$P(\gamma_j^i = 1) = 1 - P(\gamma_j^i = 0) = q$$

$$\text{for each } 1 \le i \le K, 0 \le j \le p_i$$

$$\tau_1^2(j) := (1 + c)\, \tau_1^2(0)\, \frac{e^{-Bj}}{c + e^{-Bj}}$$

$$B \quad U(0,1) \quad \text{and} \quad c \quad U(0,5)$$

$$\alpha_i \quad N(0, \sigma_\varepsilon^2 \tau_\alpha^2) \quad \text{and} \quad \sigma_\varepsilon^2 \quad \mathrm{IG}(a_\varepsilon, b_\varepsilon/2), \tag{4}$$

with $\tau_\alpha^2, \tau_0^2, \tau_1^2(0), a_\varepsilon, b_\varepsilon$ and q are hyper-parameters.

Posterior Conditional Distributions - I

The full posterior conditional distributions of the main model parameters are available in closed form, and given by

$$\theta_j^i \mid \boldsymbol{\theta}_{-i,j}, \boldsymbol{\gamma}, \sigma_\varepsilon^2, \alpha, Y \quad N\left(\mu_j^i, (\sigma_i^j)^2\right)$$

$$\gamma_j^i \mid \boldsymbol{\theta}, \sigma_\varepsilon^2, \alpha, Y \quad \text{Ber}\left(\frac{q\varphi(\theta_j^i, 0, \tau_1^2\sigma_\varepsilon^2)}{q\varphi(\theta_j^i, 0\tau_1^2\sigma_\varepsilon^2) + (1-q)\varphi(\theta_j^i, 0, \tau_0^2\sigma_\varepsilon^2)}\right),$$

$$\text{for } 1 \le i \le K, 1 \le j \le p_i$$

$$\alpha_1 \mid \boldsymbol{\theta}, \boldsymbol{\gamma}, Y \quad N\left(\mu_\alpha, \sigma_\alpha^2\right)$$

$$\sigma_\varepsilon^2 \mid \boldsymbol{\theta}, \boldsymbol{\gamma}, \alpha, Y \quad \text{IG}\left(\frac{n+p+1}{2} + a, \frac{\boldsymbol{y} - \mathbf{X}\Psi \;{}_2^2 + \boldsymbol{\theta}\ \mathbf{D}_\gamma \boldsymbol{\theta} + b}{2}\right),$$

wherein $\boldsymbol{\gamma}^i$ is the $(p_i + 1)$-dimensional vector of activity indicators corresponding to the i^{th} predictor, $\boldsymbol{\gamma} = (\boldsymbol{\gamma}^1, \boldsymbol{\gamma}^2, \cdots, \boldsymbol{\gamma}^K)$, $\mathbf{D}_\gamma = \text{Diag}\ (\tau^2_{\gamma_j^i})$, $\varphi(x, 0, \tau^2)$ is the probability density function of the normal distribution with mean zero and variance τ^2 evaluated at x; and $\boldsymbol{\theta}_{-i,j}$ is the vector obtained after omitting the entry θ_j^i from $\boldsymbol{\theta}$.

Posterior Conditional Distributions - II

These full conditional posterior distributions are easy to sample from

The parameters c and B are univariate and have a bounded range $(0, 1)$ and $(0, 5)$, respectively.
Hence, a simple accept-reject algorithm is used to generate draws from the respective conditional posterior distributions.

Numerical Illustration

- True dimension is $p = 4$ with $\beta_0 = (10.6, 7.4, 5.3, 1.5)$ and SNR $= \frac{\min \quad \beta_{0i}}{\sigma_0} = 2$.
- True model

$$y_t = (x_{t:1} \quad x_{t:2} \quad x_{t:3} \quad x_{t:4})_{4\times 1} \begin{pmatrix} 10.6 \\ 7.4 \\ 5.3 \\ 1.5 \end{pmatrix}_{4\times 1} + \varepsilon_t, \quad \text{where } \varepsilon_t \quad N(0, \sigma_0 = 0.75).$$

- $\{x_{t:k}\}$ are iid, being generated from a $N(0, 1)$ distribution.
- Run the BNL Gibbs sampler with fairly large number of parameters $p = 20$ and sample size $n = 40$.
- The average of the γ chain, estimated lag and average of the β chain are:

$$\hat{\gamma}_{NL} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 0.998 \\ \mathbf{0.034} \\ \vdots \end{pmatrix} \qquad \hat{p} = 4 \qquad \hat{\beta}_{NL} = \begin{pmatrix} 10.23 \\ 7.41 \\ 5.09 \\ 1.22 \end{pmatrix}.$$

Comparison with the 'usual' Spike-and-Slab Prior

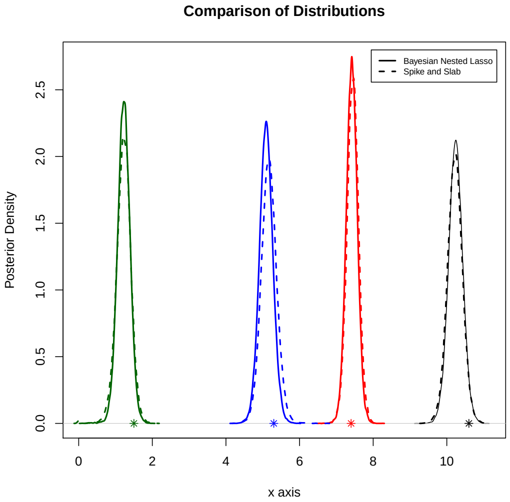
- We put independent spike-and-slab prior on each β_i (instead of on θ_i) i.e.

$$\text{For each } 1 \le i \le p \quad \begin{aligned} &\beta_i|(\sigma_\varepsilon^2, \gamma_i = 1) \quad N\left(0, \sigma_\varepsilon^2\tau_1^2\right) \\ &\beta_i|(\sigma_\varepsilon^2, \gamma_i = 0) \quad N\left(0, \sigma_\varepsilon^2\tau_0^2\right) \\ &P(\gamma_i = 1) = 1 - P(\gamma_i = 0) = q \\ &\sigma_\varepsilon^2 \quad \mathrm{IG}(a, b/2). \end{aligned}$$

- The average of the γ chain is $\hat{\gamma}_{ss}$ chain - $(1 \quad 1 \quad 1 \quad 0.98 \quad 0.082 \cdots \mathbf{0.65} \cdots 0.081)$.
- In the 15-th position the $\hat{\gamma}$ is significantly larger than 0.5 and hence the estimated lag is selected as 15.
- Relative estimation errors $\left(\left\|\hat{\beta} - \beta_0\right\| / \quad \beta_0 \quad\right)$ for $\hat{\beta}$ are given by

$$BNL = 0.035 \qquad \text{and} \qquad S\&S = 0.098.$$

Posterior Densities



Performance Evaluation - Specification of Competitors

- Standard MIDAS Ghysels et al. (2007) with Almon lag polynomials constraints on the regression coefficients
- UMIDAS-L (Unrestricted MIDAS Lasso) Uematsu and Tanaka (2019) which essentially couples unrestricted MIDAS regressions (no lag polynomial constraints on coefficients) with a standard lasso penalty on the regression coefficients.
- Bayesian BMIDAS-SS Mogliani and Simoni (2020) which couples unrestricted MIDAS regressions with a spike-and-slab group lasso prior.

Simulation Setup

(1) Data generative mechansim

$$y_t = \alpha y_{t-1} + \sum_{j=0}^{d_1} \beta_j x_{t-j/4} + \sum_{\ell=0}^{d_2} \delta_\ell z_{t-\ell/12} + \varepsilon_t, \quad (5)$$

with y being the low (yearly) frequency response variable, and x, z are high frequency predictors with 4 (quarterly) and 12 (monthly) observations, respectively for a single observation of y.

(2) x, z, ε standard normal

(3) True lags of high frequency variables x and z
$(d_1, d_2) = \{(3, 7), (5, 17), (7, 31)\}$.

(4) For each choice of (d_1, d_2), the true values of the parameters $\{\beta_0, \cdots, \beta_{d_1}\}$ and $\{\delta_0, \cdots, \delta_{d_2}\}$ are chosen in 3 different ways (described in Settings 1,2,3 below), and the true value of the single autoregressive coefficient (α) is always taken to be 0.1.

Simulation Scenarios - I

1. Neutral Setting:
   The regression coefficients are generated randomly according to a Dirichlet distribution as follows:

$$\beta_j = C_\beta \times U_j \;\text{ and }\; \delta_\ell = C_\delta \times V_\ell \quad j = 0, \cdots, d_1, \quad = 0, \cdots, d_2, \qquad (6)$$

where

$$\begin{aligned} C_\beta & \quad \text{Gamma } (8, 0.5) \text{ and } C_\delta \quad \text{Gamma } (10, 0.5) \\ (U_0, \cdots, U_{d_1}) & \quad \text{Dirichlet } (\alpha_0, \cdots, \alpha_{d_1}) \\ (V_0, \cdots, V_{d_2}) & \quad \text{Dirichlet } (\alpha_0, \cdots, \alpha_{d_2}). \end{aligned} \qquad (7)$$

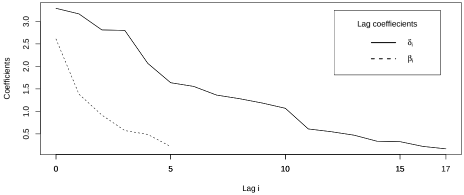


Figure: True value of β_i's (dotted) and δ_ℓ (bold) generated according to Dirichlet based procedure described in (6) and (7)

Simulation Scenarios - II

1. MIDAS Favoring Setting:
   Regression coefficients are generated according to an exponential Almon lag model. In particular, the true parameters β and δ in (5) are given by

$$\beta_j = c_\beta \frac{\exp\left(\sum_{s=1}^{2} \gamma_s j^s\right)}{\sum_{k=0}^{d_1} \exp\left(\sum_{s=1}^{2} \gamma_s k^s\right)} \qquad j = 0, 1, \cdots, d_1,$$

$$\delta_\ell = c_\delta \frac{\exp\left(\sum_{s=1}^{3} \eta_s \ {}^{s}\right)}{\sum_{k=0}^{d_2} \exp\left(\sum_{s=1}^{3} \eta_s k^s\right)} \qquad = 0, 1, \cdots, d_2. \tag{8}$$

   where $(c_\beta,\ c_\delta) = (4.5, 3.5), \quad \gamma = (1, -0.5) \quad$ and $\eta = (3, 0.2, -0.04)$. We again set the single autoregressive coefficient $\alpha = 0.1$ and consider three different choices for (d_1, d_2), namely $(3, 7)$, $(5, 17)$, $(7, 31)$.

Behavior of β and δ for $d_1 = 5$, $d_2 = 17$

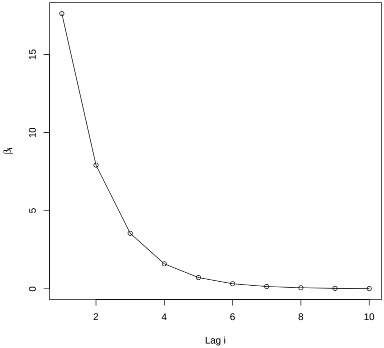


Figure: True value of β_i's (dotted) and δ_ℓ (bold) generated according to Exponential Almon Lag

Results - Neutral Setting

Relative estimation error average over 100 replications for BNL and MIDAS when the true lag coefficients are generated according to the Dirichlet based procedure in (6) and (7); (d_1, d_2) denote the true lags in the data generative model (5), and $n_1 = 15(d_2 + 1)$, $n_2 = 20(d_2 + 1)$ and $n_3 = 25(d_2 + 1)$ the sample sizes. The smallest error in each row is highlighted in **bold**.

	(d_1, d_2)	BNL	MIDAS	BMIDAS-SS	UMIDAS-L
Dirichlet procedure	$(3, 7)$	**0.69**	0.72	0.70	0.70
with $\alpha_i \equiv 1$	$(5, 17)$	**0.81**	0.88	0.82	0.82
Sample size n_1	$(7, 31)$	1.21	1.35	1.23	**1.20**
Dirichlet procedure	$(3, 7)$	**0.66**	0.69	0.68	0.69
with $\alpha_i = 1 - 1/i$	$(5, 17)$	**0.76**	0.81	0.77	0.78
Sample size n_1	$(7, 31)$	**0.93**	0.99	0.95	0.94
Dirichlet procedure	$(3, 7)$	**0.66**	0.67	0.67	0.67
with $\alpha_i \equiv 1$	$(5, 17)$	**0.81**	0.88	0.82	0.82
Sample size n_2	$(7, 31)$	**1.10**	1.32	1.13	1.14
Dirichlet procedure	$(3, 7)$	**0.65**	0.67	0.68	0.66
with $\alpha_i = 1 - 1/i$	$(5, 17)$	**0.76**	0.79	**0.76**	0.77
Sample size n_2	$(7, 31)$	**0.89**	0.96	0.91	0.90
Dirichlet procedure	$(3, 7)$	**0.63**	0.66	**0.63**	0.64
with $\alpha_i \equiv 1$	$(5, 17)$	**0.78**	0.81	0.80	0.81
Sample size n_3	$(7, 31)$	**0.99**	1.09	1.01	1.09
Dirichlet procedure	$(3, 7)$	**0.63**	0.66	0.64	0.66
with $\alpha_i = 1 - 1/i$	$(5, 17)$	**0.74**	0.78	0.75	0.77
Sample size n_3	$(7, 31)$	**0.88**	0.91	0.89	0.89

Selected Posterior Densities

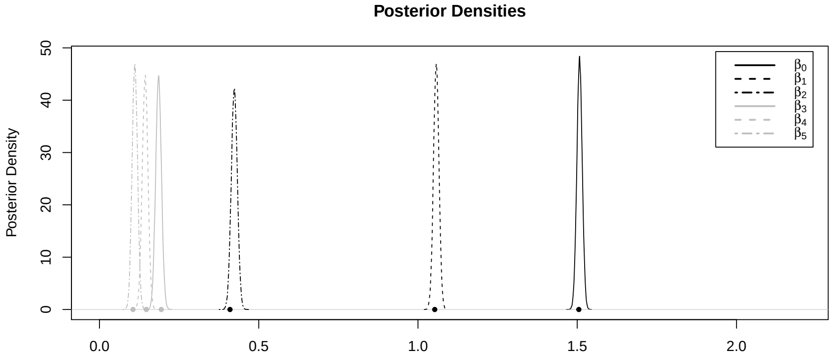


Figure: Estimated posterior densities of $\beta_0, \cdots, \beta_5$ for a replicate of the Dirichlet based setting with $d_1 = 5, d_2 = 17, n_1 = 15(d_2 + 1)$ and $\alpha_i = 1$ for every $i \geq 0$. The true values are marked on the x-axis.

Results - MIDAS Favoring Setting

Relative estimation error averaged over 100 replications for BNL and MIDAS when the true lag coefficients are generated according to exponential Almon lag polynomials described in (8) (top half) and beta Almon lag polynomials described (bottom half). **The MIDAS algorithm is supplied with the correct family and lag information**. Here (d_1, d_2) denote the true lags in the data generative model (5), and $n_1 = 15(d_2 + 1)$, $n_2 = 20(d_2 + 1)$ and $n_3 = 25(d_2 + 1)$ denote the sample sizes. The smallest error in each row is highlighted in **bold**.

	(d_1, d_2)	BNL	MIDAS	BMIDAS-SS	UMIDAS-L
Exp. Almon	$(3, 7)$	**0.59**	**0.59**	**0.59**	**0.59**
	$(5, 17)$	**0.71**	**0.71**	0.72	0.73
Sample size n_1	$(7, 31)$	1.11	**1.03**	1.10	1.05
Beta Almon	$(3, 7)$	0.66	**0.64**	0.65	0.66
	$(5, 17)$	0.76	**0.75**	0.78	0.76
Sample size n_1	$(7, 31)$	0.93	**0.91**	0.94	0.93
Exp. Almon	$(3, 7)$	0.56	**0.54**	0.57	0.55
	$(5, 17)$	0.71	**0.68**	0.71	**0.68**
Sample size n_2	$(7, 31)$	1.01	**0.97**	1.01	0.99
Beta Almon	$(3, 7)$	0.65	**0.62**	0.65	0.63
	$(5, 17)$	0.76	**0.74**	0.76	**0.74**
Sample size n_2	$(7, 31)$	0.89	**0.87**	0.89	0.89
Exp. Almon	$(3, 7)$	0.53	**0.51**	0.53	0.53
	$(5, 17)$	0.68	**0.65**	0.69	0.66
Sample size n_3	$(7, 31)$	0.89	**0.79**	0.89	**0.79**
Beta Almon	$(3, 7)$	0.63	**0.59**	0.64	0.60
	$(5, 17)$	0.74	**0.71**	0.74	0.73
Sample size n_3	$(7, 31)$	0.88	**0.85**	0.89	0.87

Selected Posterior Densities

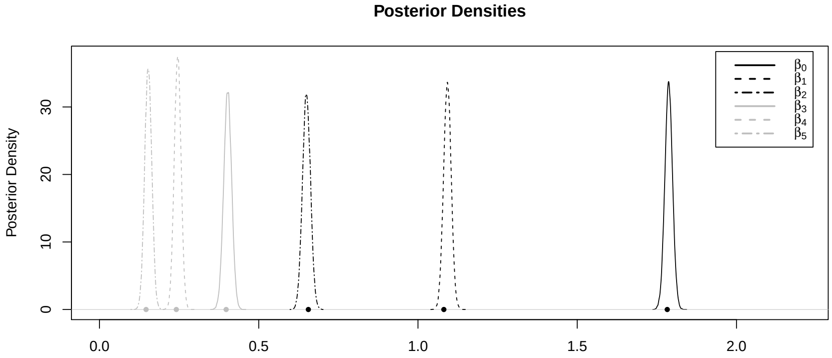


Figure: Estimated posterior densities of $\beta_0, \cdots, \beta_5$ for a replicate of the exponential Almon lag based setting in (8) with $d_1 = 5, d_2 = 17, n_1 = 15(d_2 + 1)$. The true values are marked on the x-axis.

Back to the Motivating Application

Objective: Forecast U.S. nominal GDP growth based on the FRED-MD data set containing $K = 125$ key macroeconomic variables (available monthly) for the period of January 1980 to December 2017.

Use the annualized quarterly growth rate of the GDP, $y_t = 4\log(Y_t/Y_{t-1})$ as the response variable, wherein Y_t denotes the U.S. seasonally adjusted nominal GDP

Apply various transformations (recommendations by the St. Louis Fed) to make the predictor variables stationary

Evaluate the out-of-sample forecasting performance of various methods over an evaluation window starting at Q1 of year 2000 and ending at Q4 of year 2017

Obtaining forecasts

For each method, forecasts are obtained for horizons $h = 0, 1/3, 2/3, 1, 4/3, 2$ as follows.

Let $T_0 = 1980Q1$ denote the earliest time point for which data are available, and $T_1 = 2000Q1$ and $T_2 = 2017Q4$ denote the first and last time points for the evaluation window.

For every $T \quad [T_1, T_2]$, we use data from T_0 to T to fit the appropriate version of the model which expresses the annualized GDP growth y_t in terms of the h-step prior history of the predictor variables, i.e., $\left\{ \left\{ x^i_{t-h-j/3} \right\}_{j=0}^{p_i} \right\}_{i=1}^{125}$, and the most recent autoregressive lag available.

Using the fitted model and data up to time T, a forecast $\hat{y}_{\mathrm{T+h}}$ is obtained.

The h-step ahead mean squared forecast error (MSFE) for the method is then computed as $\sum_{T \quad [T_1, T_2], T+h \leq T_2} (y_{\mathrm{T+h}} - \hat{y}_{\mathrm{T+h}})^2$.

Note that $h = 0, 1/3, 2/3$ correspond to nowcasting in the sense that we forecast quarterly (annualized) GDP growth using monthly series for very short horizons (often before the official announcement of the GDP). On the other hand $h = 1, 4/3, 2$ represent traditional short term forecast horizons.

Forecasting Results

Competitor: Naive random walk model

Relative mean squared forecasting errors for different methods and for different forecasting horizons h, with the naive Random Walk model as the benchmark. The bold values denote the best forecasting performance for each horizon.

	$h = 0$	$h = 1/3$	$h = 2/3$	$h = 1$	$h = 4/3$	$h = 2$
BNL	**0.542**	**0.586**	**0.562**	**0.614**	**0.711**	**0.763**
BMIDAS-AGL	0.661	0.681	0.689	0.812	0.894	0.932
BMIDAS-SS	0.590	0.618	0.660	0.715	0.745	0.830
UMIDAS-Lasso	0.691	0.698	0.771	0.850	0.846	0.931
UMIDAS-SCAD	0.688	0.681	0.768	0.814	0.845	0.932
UMIDAS-MCP	0.681	0.618	0.771	0.815	0.796	0.931

Some Remarks

Bayesian methods tend to outperform the MIDAS approach enhanced with regularization
Mogliani and Simoni (2020) report a similar overall conclusion

To ensure that the difference in forecasts between competing methods is not due to sample uncertainty, we used the Diebold, Mariano and West test

The null hypothesis of that test assumes unconditional equal predictive accuracy between competing methods

Almost all comparisons rejected the null hypothesis

Some BNL Model Output

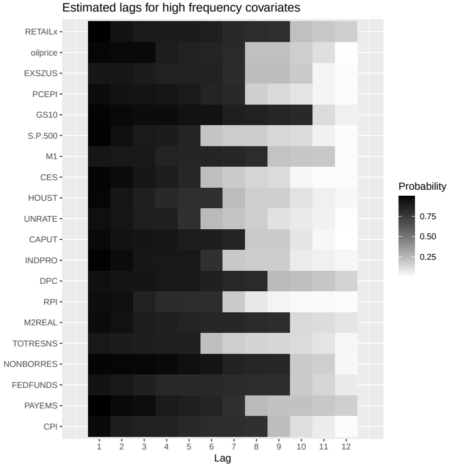


Figure: Automated lag selection: Heatmap of marginal inclusion probabilities for θ_j^i for the first 12 lags corresponding to twenty key variables for U.S. GDP forecasting. For each variable, we look for the smallest lag for which the inclusion probability is less than 0.5, and use the previous lag value as the selected lag for that variable.

Some Remarks

For most indicators the selected lag is around 6 months

For monetary (Fed Funds rate (FEDFUNDS), M1 and M2 measures of money supply (M1, M2REAL), 10 year constant maturity rate (GS10), reserves of depository institutions (NONBORRES)) and price indicators (consumer price index (CPI)) the lag increases to 9 months, indicating longer memory for forecasting purposes (**?**).

What about theoretical properties of BNL?

Under certain regularity conditions and following broadly the roadmap in Ghosh, Khare and Michailidis (2021), we can establish strong posterior consistency

Concluding Remarks

- Introduced Bayesian methodology for handling mixed frequency data, combining ideas from regularization and MIDAS
- Extensive numerical illustration shows that BNL identifies correctly the total number of parameters in the model and performs competitively vis-a-vis the MIDAS approach
- Extension: instead of using fixed values in the ‘generalized logistic function’, put a prior distribution on c, B and $\tau_1(0)$.

References I

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

Levina, E., A. Rothman, and J. Zhu (2008, 03). Sparse estimation of large covariance matrices via a nested lasso penalty. *Ann. Appl. Stat. 2*(1), 245–263.

Mogliani, M. and A. Simoni (2020). Bayesian midas penalized regressions: Estimation, selection, and prediction. *arXiv:1903.08025*.

Uematsu, Y. and S. Tanaka (2019). High-dimensional macroeconomic forecasting and variable selection via penalized regression. *The Econometrics Journal, 22 (1):34–56, 2019 22*(1), 34–56.