Variable selection in MIDAS models

A comparison of Ridge, Elastic Net, LASSO, sparse-group LASSO, OCMT, and BMT

Master thesis

Jonas Vammen Jensen

Aalborg University
Department of Mathematical Sciences
Thomas Manns Vej 23
DK-9220 Aalborg Øst

Typeset by the authors in LaTeX. Illustrations made Geogebra and R.

AALBORG UNIVERSITY
STUDENT REPORT

Fifth year w/
Department of Mathematical Sciences
Mathematics-Economics
Thomas Manns Vej 23
9220 Aalborg Øst
http://math.aau.dk

Title
Variable selection in MIDAS models: A comparison of Ridge, Elastic Net, LASSO, sparse-group LASSO, OCMT, and BMT

Theme
Macroeconomics
Finance
Machine learning

Project Period
Spring Semester 2026

Project Group
Group MO10-10

Participant
Jonas Vammen Jensen

Supervisor
Charisios Grivas

Page Count: 52
Date of Completion: May 27, 2026

Abstract

This thesis compares regularization and variable-selection methods for high-dimensional MIDAS models used to nowcast U.S. real GDP growth. The empirical analysis combines macroeconomic, financial, and textual news variables observed at different frequencies. The methods considered are Ridge, LASSO, Elastic Net, sparse-group LASSO, OCMT, and BMT.

Forecasts are produced in a rolling pseudo-out-of-sample framework from 2002Q1 to 2017Q2 at three nowcasting horizons: 2-month, 1-month, and end-of-quarter. Forecast accuracy is evaluated using RMSE, MSFE, MAE, and the aSPA test.

The results show that Ridge-U-MIDAS and sparse-group LASSO-MIDAS have the best forecasting performance, while BMT-U-MIDAS and Elastic Net-U-MIDAS also remain competitive. In contrast, OCMT-U-MIDAS has the weakest out-of-sample performance despite often selecting many predictors. Overall, the findings suggest that shrinkage and structured regularization are effective for GDP nowcasting with high-dimensional mixed-frequency data.

Preface

This project is written in the spring semester of 2026 by group MO10-10, 10th-semester students of Mathematics-Economics at the Department of Mathematical Sciences, Aalborg University, Aalborg, Denmark. The document is typeset with LaTeX.

The group would like to acknowledge the project advisor, Assistant Professor Charisios Grivas, and thank him for his encouragement, insightful comments, and guidance throughout the project period.

The following programs were used in the writing of this project:

- **Overleaf** — Writing.
- **R-studio** — Simulation and Figures.

The project has also used the AI language model ChatGPT to assist in improving the code.

Reader's Guide

The table of contents is given on page v. When the document is viewed as a PDF, entries in the table of contents act as hyperlinks, allowing simple and efficient navigation throughout the project.

The literature used in this project is introduced in the bibliography on page 43. The references in the bibliography follow the APA method shown below:

[Author][Year][Title](Publisher)(Edition)(URL).

Fields in [square brackets] are mandatory, while fields in (regular parentheses) are only relevant for certain formats (e.g. scientific articles and web pages).

Entries in the bibliography are sorted in alphabetical order. Source references are typically given at the beginning of chapters or sections. Appendices can be found following the bibliography. All programming in the project is done in R; the source code can be accessed through the following link:
`https://github.com/jonas111996/Master_Thesis`

Contents

1 | Introduction

Gross domestic product (GDP) is one of the most important indicators of macroeconomic activity. However, GDP is only observed at a quarterly frequency and is released with a delay. This creates a challenge for policymakers, financial institutions, and researchers who need timely information on the current state of the economy. In contrast, many relevant indicators are available at higher frequencies. Macroeconomic variables are often observed monthly, financial variables may be observed daily, and textual news indicators can provide continuously updated information about economic conditions. A useful objective is therefore to construct forecasting models that combine information observed at different frequencies to produce accurate short-term forecasts of GDP growth.

This project focuses on forecasting U.S. GDP growth using mixed-frequency data and variable selection methods. The analysis is based on Mixed Data Sampling (MIDAS) models. MIDAS models are used because they allow low-frequency variables, such as quarterly GDP growth, to be modeled using predictors observed at higher frequencies. This avoids first aggregating all variables to the same frequency. The predictor set consists of macroeconomic, financial, and textual news variables. The textual variables are included to examine whether economic news contains useful information for forecasting GDP growth.

A central challenge in this setting is often that the number of predictors can become large relative to the number of observations. For this reason, regularization and variable-selection methods are used. These methods reduce model complexity by shrinking coefficients or by selecting only a subset of relevant predictors. The methods considered in this project are Ridge regression, LASSO, Elastic Net, sparse-group LASSO, One Covariate Multiple Testing (OCMT), and Boosting Multiple Testing (BMT).

The predictive performance of the different models is assessed using a rolling-window approach. Forecasts are produced at three different horizons. A 2-month horizon, a 1-month horizon, and an end-of-quarter horizon. This makes it possible to examine whether the methods' accuracy changes as more recent information becomes available. The predictive performance of the models is evaluated using the root mean squared error, mean squared forecast error, mean absolute error, and the aSPA test.

2 | Mixed Data Sampling

This chapter is based on [7],[8],[1] and [10]

This chapter introduces the theoretical framework behind the MIDAS models used in this project. Hence, the definition of a MIDAS model is given as follows

Definition 1. MIDAS model
Let $\{y_t\}_{t\in\mathbb{Z}}$ be a low-frequency variable, and for each $r = 1, \ldots, R$, let $\{x_\tau^{(r)}\}_{\tau\in\mathbb{Z}}$ be a high-frequency series, observed m_r times within each low-frequency period. A MIDAS model is given by

$$y_t = \beta_0 + \sum_{r=1}^{R} \beta_r \sum_{k=0}^{K_r} w_r(k;\theta_r)\, x_{m_r t-k}^{(r)} + \epsilon_t, \qquad t \in \mathbb{Z}. \tag{2.1}$$

Here K_r denotes the number of high-frequency lags for regressor r, and

$$w_r(\cdot;\theta_r) : \{0, 1, \ldots, K_r\} \to \mathbb{R}, \qquad r = 1, \ldots, R,$$

is a weight function satisfying $\sum_{k=0}^{K_r} w_r(k;\theta_r) = 1$.

The weight functions in Definition 1 are often parameterized parsimoniously. For predictor r, let

$$u_k = \frac{k+1}{K_r+2}, \qquad k = 0, 1, \ldots, K_r,$$

so that $u_k \in (0,1)$. A common choice for weight functions in MIDAS models includes the Beta lag function, defined by

$$w_r(k;\theta_{r,1},\theta_{r,2}) = \frac{u_k^{\theta_{r,1}-1}(1-u_k)^{\theta_{r,2}-1}}{\sum_{j=0}^{K_r} u_j^{\theta_{r,1}-1}(1-u_j)^{\theta_{r,2}-1}}, \qquad \theta_{r,1},\theta_{r,2} > 0.$$

This parametrization guarantees that the lag weights are non-negative and sum to one. Depending on the values of $\theta_{r,1}$ and $\theta_{r,2}$, the resulting lag weights can be decreasing, hump-shaped, or concentrated on recent or distant lags. Another widely used choice is the Exponential Almon lag function.

$$w_r(k;\theta_r) = \frac{\exp\big(\theta_{r,1}k + \theta_{r,2}k^2 + \cdots + \theta_{r,Q}k^Q\big)}{\sum_{j=0}^{K_r} \exp\big(\theta_{r,1}j + \theta_{r,2}j^2 + \cdots + \theta_{r,Q}j^Q\big)}.$$

This parametrization likewise guarantees that the lag weights sum to one. Alternatively, the lag weights can be approximated using basis expansions such as Legendre polynomials.

These orthogonal polynomials are often attractive because they reduce multicollinearity and typically have better numerical properties than raw Almon polynomials. Let

$$\phi_\ell(u) = P_\ell(2u-1), \qquad u \in [0,1],$$

where P_ℓ denotes the ℓ-th Legendre polynomial on $[-1,1]$. The unrestricted lags can be approximated as

$$\tilde{w}_r(k;\theta_r) = \sum_{\ell=0}^{L_r} \theta_{r,\ell}\phi_\ell(u_k), \qquad k = 0,\dots,K_r.$$

Since this expansion does not, in general, impose non-negativity or the normalization $\sum_{k=0}^{K_r} \tilde{w}_r(k;\theta_r) = 1$, it should be interpreted as an approximation to the lag pattern rather than as a normalized MIDAS weight function. In a standard MIDAS model, the coefficients on the high-frequency lags are constrained to follow such a parsimonious weight function, whereas in a U-MIDAS model, these lag coefficients are left unrestricted and estimated directly. The model is called a single-regressor MIDAS model when only one mixed-frequency regressor is included and a multiple-regressor MIDAS model when several mixed-frequency regressors are included.

Sometimes, the dependent variable depends not only on current high-frequency information, but also on its own previous values. In such cases, the MIDAS model can be extended to include lagged values of the dependent variable. This leads to an autoregressive distributed lag MIDAS model, often called an ARDL-MIDAS model.

Definition 2. ARDL-MIDAS model
Let $\{y_t\}_{t\in\mathbb{Z}}$ be a low-frequency variable, and for each $r = 1,\dots,R$, let $\{x_\tau^{(r)}\}_{\tau\in\mathbb{Z}}$ be a high-frequency series, observed m_r times within each low-frequency period. Then, an ARDL-MIDAS model can be written as

$$y_t = \beta_0 + \sum_{j=1}^{P} \rho_j y_{t-j} + \sum_{r=1}^{R} \beta_r \sum_{k=0}^{K_r} w_r(k;\theta_r)\, x_{m_r t-k}^{(r)} + \epsilon_t, \qquad t \in \mathbb{Z}. \tag{2.2}$$

Here, P denotes the number of low-frequency lags of the dependent variable, K_r is the number of high-frequency lags for regressor r, and

$$w_r(\cdot;\theta_r) : \{0,1,\dots,K_r\} \to \mathbb{R}, \qquad r = 1,\dots,R,$$

is a weight function for regressor r satisfying $\sum_{k=0}^{K_r} w_r(k;\theta_r) = 1$.

As previously mentioned, this model is also referred to as an ARDL-U-MIDAS model if the lag coefficients are left unrestricted. In the next section, we discuss how the parameters in (U)MIDAS and ARDL-MIDAS models can be estimated.

2.1 Estimation of MIDAS models

In this section, different estimation approaches for MIDAS models are introduced. It is useful to first note that U-MIDAS models can be estimated by OLS, since the lag coefficients are left unrestricted and the model is therefore linear in the parameters. For restricted MIDAS models, where the lag coefficients are parameterized through weighting functions, other estimation methods are typically needed. We consider a class of extremum estimators, where γ denotes the finite-dimensional vector of unknown parameters. An estimator $\hat{\gamma}_T$ is called an extremum estimator if there exists an objective function $\hat{M}_T(\gamma)$ such that $\hat{\gamma}_T \in \arg\max_{\gamma\in\Gamma} \hat{M}_T(\gamma)$. In particular, the objective function for MLE is the average log-likelihood.

$$\hat{M}_T(\gamma) = \frac{1}{T}\sum_{t=1}^{T} l(\epsilon_t|\gamma)$$

where l denotes the log-likelihood contribution implied by the assumed distribution of $\epsilon_t(\gamma)$, where $\epsilon_t(\gamma)$ is the error term from (2.1) or (2.2). However, the objective function of the NLS estimator is defined as follows.

$$\hat{M}_T(\gamma) = -\frac{1}{T}\sum_{t=1}^{T} \epsilon_t(\gamma)^2.$$

Here, ϵ_t denotes the error term obtained by isolating in either (2.1) or (2.2), depending on whether the model under consideration is a standard MIDAS model or an ARDL-MIDAS model. Finally, the GMM estimator for the objective function is given as follows.

$$\hat{M}_T(\gamma) = -\left[\frac{1}{T}\sum_{t=1}^{T} g_t(\gamma)\right]^{\top} \hat{W}_T \left[\frac{1}{T}\sum_{t=1}^{T} g_t(\gamma)\right]$$

where $g_t(\gamma) = \epsilon_t(\gamma)Z_{t-1}$ and Z_{t-1} is an instrument vector.

One of the usual regularity conditions required for consistency of the estimators is compactness of the parameter space. This is commonly imposed by assuming that the parameter space is finite-dimensional as well as closed and bounded. Specifically, let $\Gamma \subset \mathbb{R}^q$, and suppose that Γ is compact. A further essential condition for consistency is identification, which can be expressed as follows.

Assumption 2.1.1. Assume that there exists a unique parameter vector $\gamma_0 \in \Gamma \subset \mathbb{R}^q$, where Γ is compact, such that the error term $\varepsilon_t(\gamma_0)$ from either the MIDAS model or the ARDL-MIDAS model satisfies

$$\mathbb{E}\big[\epsilon_t(\gamma_0) \mid \mathcal{I}_t\big] = 0.$$

Here $\mathcal{I}_t$ denotes the information set generated by the observed high-frequency regressors up to time t, and, in the ARDL-MIDAS case, also by lagged values of the low-frequency variable y_t.

This assumption imposes a conditional mean restriction and a uniqueness condition for the true parameter. Together with standard regularity conditions, such as compactness of the parameter space, continuity of the criterion function, and uniform convergence of the sample criterion, it can be used to establish consistency of the estimator.

Having established how the parameters of (U-)MIDAS and ARDL-(U)MIDAS models can be estimated under suitable regularity conditions, we now turn to methods for evaluating the predictive performance of these models, which are introduced in the next section

2.2 Average Superior Predictive Ability test

In this section, we introduce a method for comparing the predictive performance of different MIDAS models across several forecasting horizons, namely the Average Superior Predictive Ability (aSPA) test. Instead of evaluating forecasting performance at each horizon separately, the aSPA test evaluates whether one model has superior predictive ability on average across all horizons.

Consider two models, denoted by model i and model j. Let $\widehat{y}_{i,t+h|t}$ and $\widehat{y}_{j,t+h|t}$ denote the forecasts of y_{t+h} made at time t for horizon $h = 1, \ldots, H$. The corresponding forecast errors are defined as

$$e_{i,t+h|t} = y_{t+h} - \widehat{y}_{i,t+h|t}, \qquad e_{j,t+h|t} = y_{t+h} - \widehat{y}_{j,t+h|t}.$$

For a given loss function $L(\cdot)$, the loss differential is defined as

$$d_{ij,t}^{(h)} = L(e_{i,t+h|t}) - L(e_{j,t+h|t}), \qquad h = 1, \ldots, H.$$

A positive value of $d_{ij,t}^{(h)}$ means that model j has a smaller loss than model i at horizon h. Next, let $\mu_{ij}^{(h)} = \mathbb{E}[d_{ij,t}^{(h)}]$ denote the expected loss differential at horizon h. Then the vector of expected loss differentials across the H horizons is $\mu_{ij} = (\mu_{ij}^{(1)}, \ldots, \mu_{ij}^{(H)})^\top$. The aSPA test compares the models using a weighted average of the expected loss differentials.

$$\mu_{ij}^{(\text{Avg})} = w^\top \mu_{ij} = \sum_{h=1}^{H} w_h \mu_{ij}^{(h)},$$

where $w = (w_1, \ldots, w_H)^\top$ is a vector of non-negative weights satisfying $\sum_{h=1}^{H} w_h = 1$. The null hypothesis of the aSPA test is

$$H_{0,\text{aSPA}} : \mu_{ij}^{(\text{Avg})} \leq 0,$$

and the alternative hypothesis is

$$H_{1,\text{aSPA}} : \mu_{ij}^{(\text{Avg})} > 0.$$

Since forecast errors and loss differentials may be serially correlated, inference must account for the long-run variance of the weighted loss differential.

$$\zeta_{ij}^2 = \sum_{k=-\infty}^{\infty} \mathrm{Cov}(s_{ij,t}, s_{ij,t-k}).$$

where $s_{ij,t} = \sum_{h=1}^{H} w_h d_{ij,t}^{(h)}$. Typically ζ_{ij}^2 is estimated using a heteroskedasticity and autocorrelation consistent estimator. An example could be the Bartlett weights, defined as follows

$$\widehat{\zeta}_{ij}^2 = \widehat{\gamma}_s(0) + 2\sum_{k=1}^{M} \left(1 - \frac{k}{M+1}\right) \widehat{\gamma}_s(k),$$

where M is the truncation lag and

$$\widehat{\gamma}_s(k) = \frac{1}{T} \sum_{t=k+1}^{T} (s_{ij,t} - \overline{s}_{ij})(s_{ij,t-k} - \overline{s}_{ij}).$$

Here $\overline{s}_{ij} = \frac{1}{T}\sum_{t=1}^{T} s_{ij,t}$.The aSPA test statistic is defined as the following.

$$t_{\mathrm{aSPA},ij} = \frac{\sqrt{T}\,\overline{s}_{ij}}{\widehat{\zeta}_{ij}}.$$

In practice, critical values are obtained using a bootstrap procedure. This is important because the loss differentials may depend on both time and horizon. A moving block bootstrap can be used to capture this dependence structure. For each bootstrap replication $b = 1, \ldots, B$, blocks of the vector-valued loss differential

$$d_{ij,t} = \left(d_{ij,t}^{(1)}, \ldots, d_{ij,t}^{(H)}\right)^{\top}$$

are resampled to obtain a bootstrap sample $d_{ij,t}^b$. The weighted average bootstrap loss differential is then computed as

$$\overline{d}_{ij}^b = \frac{1}{T}\sum_{t=1}^{T} w^{\top} d_{ij,t}^b.$$

The corresponding bootstrap statistic is

$$t_{\mathrm{aSPA},ij}^b = \frac{\sqrt{T}\left(\overline{d}_{ij}^b - \overline{d}_{ij}\right)}{\widehat{\zeta}_{ij}^b}.$$

The p-value is computed as the fraction of bootstrap statistics that exceed the observed test statistic.

$$p_{\mathrm{aSPA}} = \frac{1}{B}\sum_{b=1}^{B} \mathbf{1}\left\{t_{\mathrm{aSPA},ij}^b \geq t_{\mathrm{aSPA},ij}\right\}.$$

If this p-value is below the chosen significance level, the null hypothesis is rejected. Thus, the test concludes that model j has average superior predictive ability relative to model i across the horizons. Having introduced the (U)-MIDAS and ARDL-(U)-MIDAS models and a method for evaluating predictive accuracy, we now turn to regularization methods. These can reduce complexity and potentially improve predictive performance.

3 | Regularization methods

This chapter is based on [1],[3],[9] and [6]

This chapter presents regularization and variable-selection methods for high-dimensional MIDAS regressions. Some methods, such as LASSO, Sparse-Group LASSO, One Covariate Multiple Testing, and Boosting Multiple Testing, can be used for variable selection, whereas ridge mainly provides coefficient shrinkage. These methods are useful when the number of potential predictors is large relative to the sample size.

3.1 LASSO, Ridge Regression and Elastic Net

Consider a linear regression, where $Y \in \mathbb{R}$ denotes the outcome variable and $X \in \mathbb{R}^p$ the vector of predictors. The conditional mean is approximated by the linear model $\mathbb{E}(Y \mid X = x) = \beta_0 + x^\top \beta$. Next, suppose we observe N realizations of (x_i, y_i) for $i = 1, \ldots, N$. For convenience, assume that the predictors are standardized such that, for each $j = 1, \ldots, p$,

$$\sum_{i=1}^{N} x_{ij} = 0 \qquad \text{and} \qquad \frac{1}{N}\sum_{i=1}^{N} x_{ij}^2 = 1.$$

This method can be extended straightforwardly to the case of unstandardized predictors. The elastic net estimator is defined as the solution to the following optimization problem.

$$\min_{(\beta_0,\beta)\in\mathbb{R}^{p+1}} \left[\frac{1}{2N} \sum_{i=1}^{N} \left(y_i - \beta_0 - x_i^\top \beta \right)^2 + \lambda P_\alpha(\beta) \right], \tag{3.1}$$

where $P_\alpha(\beta)$ is defined as

$$P_\alpha(\beta) = (1-\alpha)\frac{1}{2}\|\beta\|_{\ell_2}^2 + \alpha\|\beta\|_{\ell_1} = \sum_{j=1}^{p} \left[\frac{1}{2}(1-\alpha)\beta_j^2 + \alpha|\beta_j| \right].$$

Here $P_\alpha(\beta)$ denotes the regularization penalty. The second equality follows from the definitions of the ℓ_1- and ℓ_2-norms. When $\alpha = 1$, it reduces to the LASSO penalty, whereas for $\alpha = 0$ it becomes the ridge penalty. This penalty is particularly useful when the number of predictors is much higher than the number of observations. However, the Elastic-Net regression is obtained by letting $\alpha = 1 - v$ for some v between 0 and 1. One way to solve (3.1) is by coordinate descent. Hence, suppose that $\tilde{\beta}_0$ and $\tilde{\beta}_{\mathcal{L}}$ for $\mathcal{L} \neq j$ are given. The goal is to partially optimize with respect to β_j. To do this, let $R(\beta_0, \beta)$ be the objective function in (3.1). The partial optimality condition for β_j is then obtained by computing the gradient at $\beta_j = \tilde{\beta}_j$. Note that this gradient exists only if $\tilde{\beta}_j \neq 0$, since the ℓ_1-penalty is not differentiable at zero. However, if $\tilde{\beta}_j > 0$, the gradient is given as

$$\left.\frac{\partial R}{\partial \beta_j}\right|_{\beta=\tilde{\beta}} = -\frac{1}{N}\sum_{i=1}^{N} x_{ij}\left(y_i - \tilde{\beta}_0 - x_i^\top \tilde{\beta}\right) + \lambda(1-\alpha)\tilde{\beta}_j + \lambda\alpha.$$

A similar expression holds when $\tilde{\beta}_j < 0$, with the last term replaced by $-\lambda\alpha$. Note that the derivative does not exist at $\tilde{\beta}_j = 0$, since the ℓ_1-penalty is not differentiable at zero. This case is handled using subgradient arguments. Since the squared-error loss, the ℓ_2-penalty, and the ℓ_1-penalty are convex, the elastic-net objective is convex. Moreover, when $\lambda(1-\alpha) > 0$, the ridge component makes the criterion strictly convex in β, implying a unique minimizer. However, the derivative does not exist at $\beta_j = 0$, so this case has to be treated separately using subgradient arguments. According to [4], the coordinate-wise update has the following form.

$$\tilde{\beta}_j \leftarrow \frac{S\left(\frac{1}{N}\sum_{i=1}^{N} x_{ij}\left(y_i - \tilde{y}_i^{(j)}\right), \lambda\alpha\right)}{1+\lambda(1-\alpha)}.$$

where $\tilde{y}_i^{(j)} = \tilde{\beta}_0 + \sum_{\ell \neq j} x_{i\ell}\tilde{\beta}_\ell$ is the fitted value with the contribution of x_{ij} removed. Therefore, $y_i - \tilde{y}_i^{(j)}$ is the partial residual used for fitting β_j. Since the predictors are standardized, $\frac{1}{N}\sum_{i=1}^{N} x_{ij}\left(y_i - \tilde{y}_i^{(j)}\right)$ becomes the least-squares coefficient obtained by regressing this partial residual on x_{ij}. Here, $S(z,\gamma)$ denotes the soft-thresholding operator, defined by

$$S(z,\gamma) = \operatorname{sign}(z)(|z|-\gamma)_+ = \begin{cases} z-\gamma, & z > \gamma, \\ z+\gamma, & z < -\gamma, \\ 0, & |z| \leq \gamma. \end{cases}$$

For each coefficient, we first regress the partial residual on the corresponding predictor to obtain the ordinary least-squares update. We then pass this value through the soft-thresholding step, which incorporates the sparsity effect of the lasso penalty by possibly setting the coefficient to zero. Finally, we scale the result downward to account for the additional shrinkage imposed by the ridge part of the penalty. In the following section, another method for reducing the number of predictors is the sparse-group LASSO. This method combines LASSO and group LASSO.

3.2 Sparse-group LASSO

This section introduces an extension of the LASSO in which the lags associated with each predictor are treated as groups. Let $y = (y_1, .., y_T)^\top$ denote the outcome variable, and define the design matrix $X = \left(\iota, y_1, \ldots, y_J, Z_1W, \ldots, Z_KW\right)$, where $\iota = (1, 1, \ldots, 1)^\top$ is a $T \times 1$ vector of ones. Furthermore let $y_j = (y_{1-j}, \dots, y_{T-j})^\top$ denote the j-th lag of the outcome variable, and $Z_k = \big(x_{k,t-(j-1)/m}\big)_{t\in[T],\, j\in[m]}$ be a $T \times m$ matrix associated with predictor $k \in [K]$, and $W = \big(w_l((j-1)/m)/m\big)_{j\in[m],\, l\in[L]}$ an $m \times L$ matrix of weights. Next, define

$$\beta = (\beta_{AR}^\top, \beta_1^\top, \dots, \beta_K^\top)^\top,$$

where $\beta_{AR} = (\beta_0, \rho_1, \dots, \rho_J)^\top$ contains the intercept and autoregressive coefficients. Each of the $\beta_k \in \mathbb{R}^L$ contains the parameters of the high-frequency lag polynomial associated with predictors $k \geq 1$. The sparse-group LASSO estimator, denoted by $\hat{\beta}$, is defined as the solution to the following penalized least-squares problem

$$\min_{b\in\mathbb{R}^p} \|y - Xb\|_T^2 + 2\lambda\Omega(b). \tag{3.2}$$

Here, the penalty function combines the LASSO penalty and the group penalty.

$$\Omega(b) = \alpha\|b\|_1 + (1-\alpha)\|b\|_{2,1}$$

where $\|b\|_{2,1} = \sum_{G\in\mathcal{G}} \|b_G\|_2$ is the group LASSO norm and $\mathcal{G}$ is a partition of $\{1, .., p\}$. The estimator in (3.2) can be computed using a coordinate descent or block-coordinate descent algorithm.

The penalty in (3.2) is determined by the parameter $\lambda > 0$, and $\alpha \in [0, 1]$ is a weight parameter, which determines the relative importance of sparsity and group structure. In the case where $\alpha = 1$, the LASSO estimator is obtained. However, if $\alpha = 0$, the group LASSO estimator is obtained. The value of α is typically determined by fixing it or selecting it jointly with λ using cross-validation. The following section introduces an alternative regularization method. Instead of estimating a single model with all predictors simultaneously, it tests each predictor's contribution one at a time.

3.3 One Covariate Multiple Testing

In this section, another variable-selection method for high-dimensional linear regression is presented, called One Covariate Multiple Testing (OCMT). As mentioned previously, this method estimates the contribution of each predictor one at a time instead of estimating the whole model at once. Let $\{y_t\}_{t=1}^T$ denote the dependent variable and let $\{x_{i,t}\}_{t=1}^T$, for $i = 1, \dots, n$, denote the set of possible predictors. In the first stage of the OCMT procedure, we run n separate bivariate regressions of y_t on a constant and each predictor $x_{i,t}$ taken one at a time. Then we obtain

$$y_t = c_{i,(1)} + \phi_{i,(1)} x_{i,t} + \epsilon_{i,t,(1)}, \qquad t = 1, 2, \dots, T,$$

where $\phi_{i,(1)} = \theta_i/\sigma_{ii}$ and θ_i is defined as follows.

$$\theta_{i,T} = \sum_{j=1}^{n} \mathbb{1}(\beta_j \neq 0)\beta_j\sigma_{ij,T} = \sum_{j=1}^{k} \beta_j\sigma_{ij,T}.$$

Here

$$\sigma_{ij,T} = \mathbb{E}\left[T^{-1}x_i^\top M_\tau x_j\right], \qquad M_\tau = I_T - \frac{\tau_T\tau_T^\top}{T},$$

where $x_i = (x_{i,1}, x_{i,2}, \ldots, x_{i,T})^\top$ and τ_T is a $T \times 1$ vector of ones. To simplify notation, the subscript T is suppressed below, so that θ_i and σ_{ij} denote $\theta_{i,T}$ and $\sigma_{ij,T}$, respectively.

For later stages of the OCMT procedure, it is useful to define a conditional net effect. Let Z denote a matrix containing the constant and a subset of already selected predictors. Define

$$M_Z = I_T - Z(Z^\top Z)^{-1}Z^\top.$$

Then the conditional covariance and conditional net effect are given by $\sigma_{ij,T}(Z) = \mathbb{E}\left[T^{-1}x_i^\top M_Z x_j\right]$ and $\theta_{i,T}(Z) = \sum_{j=1}^k \beta_j\sigma_{ij,T}(Z)$. Now let the t-statistic of $\phi_{i,(1)}$ be defined as

$$t_{\hat{\phi}_{i,(1)}} = \frac{\hat{\phi}_{T,i,(1)}}{\text{s.e.}\left(\hat{\phi}_{T,i,(1)}\right)} = \frac{x_i^\top M_{(0)}y}{\hat{\sigma}_{i,(1)}\sqrt{x_i^\top M_{(0)}x_i}},$$

where $y = (y_1, y_2, \ldots, y_T)^\top$ denotes the $T \times 1$ vector of observations on the dependent variable. Moreover, $\hat{\phi}_{T,i,(1)}$ is the OLS estimator from the regression of y on a constant and x_i,

$$\hat{\sigma}^2_{i,(1)} = \frac{\epsilon_{i,(1)}^\top\epsilon_{i,(1)}}{T}, \qquad \epsilon_{i,(1)} = M_{i,(0)}y,$$

with

$$X_{i,(0)} = (x_i, \tau_T), \qquad M_{i,(0)} = I_T - X_{i,(0)}(X_{i,(0)}^\top X_{i,(0)})^{-1}X_{i,(0)}^\top, \qquad M_{(0)} = I_T - \frac{\tau_T\tau_T^\top}{T}.$$

The first-stage multiple-testing indicator for whether predictor i is selected is given by

$$\mathbb{1}_{(1)}(\beta_i \neq 0) = \mathbb{1}\left(|t_{\hat{\phi}_{i,(1)}}| > c_p(n,\delta)\right), \qquad i = 1, 2, \ldots, n,$$

where $c_p(n,\delta)$ is the critical value given by

$$c_p(n,\delta) = \Phi^{-1}\left(1 - \frac{p}{2f(n,\delta)}\right). \tag{3.3}$$

Here $\Phi^{-1}(\cdot)$ is the inverse standard normal cdf, $f(n,\delta) = cn^\delta$ for some positive constants δ, c, and $p \in (0,1)$. The parameter p denotes the nominal size of the individual test, while δ is called the critical value exponent. Predictors for which the multiple-testing indicator equals 1 are selected in the first stage.

Next, in the second stage, let $\hat{k}^o_{n,T,(1)}$ be the number of predictors selected in the first stage, let $S^o_{(1)}$ be the index set of the selected predictors, and let $X^o_{(1)}$ denote the $T \times \hat{k}^o_{n,T,(1)}$ design matrix of the selected predictors. Now, let

$$X_{(1)} = (\tau_T, X^o_{(1)}) = (x_{(1),1}, \ldots, x_{(1),T})^\top, \qquad \hat{k}_{n,T,(1)} = \hat{k}^o_{n,T,(1)},$$

and define

$$S_{(1)} = S^o_{(1)}, \qquad N_{(1)} = \{1, 2, \ldots, n\} \setminus S_{(1)}.$$

For stages $j = 2, 3, \ldots$, we run $n - \hat{k}_{n,T,(j-1)}$ regressions of y_t on the variables contained in $X_{(j-1)}$, adding each $x_{i,t}$ separately for $i \in N_{(j-1)}$. The corresponding t-statistic is then given by

$$t_{\hat{\phi}_{T,i,(j)}} = \frac{\hat{\phi}_{T,i,(j)}}{\text{s.e.}\left(\hat{\phi}_{T,i,(j)}\right)} = \frac{x_i^\top M_{(j-1)} y}{\hat{\sigma}_{i,(j)} \sqrt{x_i^\top M_{(j-1)} x_i}}, \qquad i \in N_{(j-1)},\ j = 2, 3, \ldots.$$

Here $\hat{\phi}_{T,i,(j)} = \hat{\phi}_{i,(j)} = \left(x_i^\top M_{(j-1)} x_i\right)^{-1} x_i^\top M_{(j-1)} y$ is the estimated partial effect of $x_{i,t}$ on y_t at stage j, $\hat{\sigma}^2_{i,(j)} = T^{-1} \epsilon^\top_{i,(j)} \epsilon_{i,(j)}$ is the corresponding estimator of the residual variance, and $M_{(j-1)} = I_T - X_{(j-1)} \left(X^\top_{(j-1)} X_{(j-1)}\right)^{-1} X^\top_{(j-1)}$ removes the influence of the predictors already contained in $X_{(j-1)}$. . Moreover, $\epsilon_{i,(j)}$ denotes the residuals from the regression of y on $X_{i,(j-1)} = (x_i, X_{(j-1)})$. At stage j, the selection indicator is given by

$$\mathbb{1}_{(j)}(\beta_i \neq 0) = \mathbb{1}\left(|t_{\hat{\phi}_{T,i,(j)}}| > c_p(n, \delta)\right), \qquad i \in N_{(j-1)}.$$

Predictors for which this selection indicator equals 1 are then added to the set of selected predictors. At stage j, let $\hat{k}^o_{n,T,(j)}$ denote the number of predictors selected at that stage, $S^o_{(j)}$ the corresponding set of selected indices, and $X^o_{(j)}$ denotes the $T \times \hat{k}^o_{n,T,(j)}$ matrix containing the selected predictors. Next, define

$$X_{(j)} = \left(X_{(j-1)}, X^o_{(j)}\right) = \left(x_{(j),1}, x_{(j),2}, \ldots, x_{(j),T}\right)^\top,$$

and update the total number of selected predictors according to $\hat{k}_{n,T,(j)} = \hat{k}_{n,T,(j-1)} + \hat{k}^o_{n,T,(j)}$. Likewise, let $S_{(j)} = S_{(j-1)} \cup S^o_{(j)}$ denote the updated index set of selected variables. Furthermore let

$$N_{(j)} = \{1, 2, \ldots, n\} \setminus S_{(j)}$$

be the set of remaining candidates. Note that $\hat{k}_{n,T,(j)}$ is the total number of predictors selected up to and including stage j. In addition,

$$\hat{\phi}_{T,i,(j)} \xrightarrow{p} \frac{\theta_i(X_{(j-1)})}{\sigma_{ii}(X_{(j-1)})},$$

where $\theta_i(X_{(j-1)})$ denotes the conditional net effect of predictor i, after projecting out the variables already selected in the previous stages. For $j = 1$, this reduces to the unconditional net effect θ_i. The procedure ends at the first stage, where no additional predictors are selected. If this stage is denoted by $\hat{\jmath}_{n,T}$, then the total number of selected variables is

$$\hat{k}_{n,T} = \hat{k}_{n,T,(\hat{\jmath}_{n,T}-1)}.$$

This completes the procedure.

3.4 Boosting Multiple Testing

This section introduces an alternative variable-selection method, referred to as Boosting with Multiple Testing (BMT). In contrast to OCMT, the BMT procedure selects only the single most significant regressor at each stage, conditional on the variables previously included. The remaining predictors are then re-tested in the next iteration. In the first stage, we estimate n separate regressions of y_t on each candidate regressor x_{it}, for $i = 1, \ldots, n$. Let $t_{i,(1)}$ denote the corresponding t-statistic, where the subscript identifies the stage of the procedure. The regressor selected in this step is the one associated with the largest absolute t-statistic. Its index is denoted by $S^c_{(1)}$, and its $T \times 1$ vector of observations is written as $x^o_{(1)}$. We then define the initial regressor matrix as

$$X^{(1)} = (Z, x^o_{(1)}),$$

where $Z = (z_1, \ldots, z_T)^\top$ contains the pre-selected predictors. The set of selected variables after the first stage is therefore

$$S_{(1)} = S^c_{(1)},$$

while the remaining candidate variables are collected in

$$A_{(2)} = \{1, \ldots, n\} \setminus S_{(1)}.$$

For any subsequent stage $j \geq 2$, we regress y_t on the variables already included in $X_{(j-1)}$ together with each remaining candidate regressor x_{it}, for all $i \in A_{(j)}$. Let $t_{i,(j)}$ denote the associated t-statistic for x_{it}. A candidate predictor is said to pass the threshold whenever

$$\hat{L}_{i,(j)} = \mathbb{1}\left(|t_{i,(j)}| > c_p(|A_{(j)}|, \delta)\right).$$

where the critical value is given in (3.3). Among the regressors that satisfy this threshold, only the one with the largest absolute t-statistic is added to the model. Denoting its index by $S^c_{(j)}$ and its $T \times 1$ observation vector by $x^o_{(j)}$, the regressor matrix is updated as

$$X_{(j)} = \left(X_{(j-1)}, x^o_{(j)}\right),$$

and the set of selected variables becomes $S_{(j)} = S_{(j-1)} \cup S^c_{(j)}$. Accordingly, the set of remaining candidate regressors for the next iteration is

$$A_{(j+1)} = \{1, \ldots, n\} \setminus S_{(j)}.$$

Thus, the BMT procedure is sequential. At each stage, the model conditions on the predictors selected in earlier stages, applies a multiple-testing rule to the variables still under consideration, and includes at most one additional regressor. The algorithm stops once no remaining candidate satisfies the required threshold. If $\hat{J}$ denotes the final stage at which a new variable is admitted, the resulting matrix is given by $X^{(\hat{J})}$. This matrix contains both the pre-selected predictors and the variables chosen by the BMT procedure.

4 | Application

We have now established the full theoretical framework for the models used to forecast U.S. GDP growth. In this chapter, we show how the concepts and models introduced earlier are implemented in practice. We begin by gathering historical data on the U.S. GDP growth rate as the low-frequency variable from the Federal Reserve Bank of St. Louis [5]. The predictors are also obtained from the Federal Reserve Bank of St. Louis [5], including quarterly and monthly data. The predictors can be found in Table 4.1.

In addition, we use historical textual variables from The Structure of Economic News [2]. This dataset contains the full text content of 800,000 Wall Street Journal articles from 1984 to 2017. These variables summarize news into different topics, such as employment, housing, finance, or consumer demand, and measure how much attention each topic receives over time. In this thesis, only the following groups are used: Banks, Economic growth, Financial markets, Government, Industry, International affairs, Labor and income, and Oil and mining. The variables contained in these groups can be found in Table 4.1 and Figure B.1.

Table 4.1: Data used in the forecasting application

Block	**Series**	**Description**	**Freq.**	**Transformation**
Target	Real GDP growth	Annualized real GDP growth	Quarterly	$100[(RGDP_t/RGDP_{t-1})^4 - 1]$
Macro	`PERMIT`	Building permits	Monthly	First difference
Macro	`TCU`	Capacity utilization	Monthly	First difference
Macro	`UNRATE`	Unemployment rate	Monthly	First difference
Macro	`CPIAUCSL`	CPI	Monthly	Percent change
Macro	`CPILFESL`	Core CPI	Monthly	Percent change
Macro	`HOUST`	Housing starts	Monthly	Percent change
Macro	`INDPRO`	Industrial production	Monthly	Percent change
Macro	`BUSINV`	Business inventories	Monthly	Percent change
Macro	`DGORDER`	Durable goods orders	Monthly	Percent change
Macro	`HSN1F`	New one-family houses sold	Monthly	Percent change
Macro	`ULCNFB`	Unit labor costs	Quarterly	Annualized percent change
Macro	`PCEPILFE`	Core PCE price index	Monthly	Percent change
Macro	`PCEPI`	PCE price index	Monthly	Percent change
Macro	`RSAFS`	Retail sales	Monthly	Percent change
Macro	`PCEC96`	Real PCE	Monthly	Percent change
Macro	`DSPIC96`	Real disposable income	Monthly	Percent change
Macro	`PAYEMS`	Nonfarm payrolls	Monthly	First difference
Macro	`JTSJOL`	Job openings	Monthly	First difference
Macro	`MANEMP`	Manufacturing employment	Monthly	First difference
Macro	`BOPTEXP`	Exports	Monthly	Percent change
Macro	`BOPTIMP`	Imports	Monthly	Percent change
Macro	`IQ`	Export price index	Monthly	Percent change
Macro	`IR`	Import price index	Monthly	Percent change
Macro	`AMDMVS`	Durable shipments	Monthly	Percent change
Macro	`AMDMTI`	Durable inventories	Monthly	Percent change
Macro	`AMTMUO`	Unfilled orders	Monthly	Percent change
Macro	`GACDINA066MNFRBNY`	NY Fed business conditions	Monthly	Level
Macro	`GACDFSA066MSFRBPHI`	Philly Fed business outlook	Monthly	Level

Block	Series	Description	Freq.	Transformation
Macro	A261RX1Q020SBEA	Real final sales	Quarterly	Annualized percent change
Macro	TTLCONS	Construction spending	Monthly	Percent change
Financial	BAA_AAA	BAA–AAA spread	Monthly	BAA minus AAA
Financial	BAA_10Y	BAA–10Y spread	Monthly	BAA minus 10Y Treasury
Financial	SP500_ret	S&P 500 return	Monthly	$100\Delta\log(SP500)$
Financial	TED	TED spread	Monthly	Level
Financial	YC_10Y_3M	Yield curve slope	Monthly	10Y minus 3M Treasury
Financial	VIX	VIX index	Monthly	Level
Financial	EPUI	Economic policy uncertainty	Monthly	Level
Financial	EMEUI	Equity market uncertainty	Monthly	Level
Textual	Banks	Bank-related news topics	Monthly	Topic attention share
Textual	Economic growth	Growth and macro news topics	Monthly	Topic attention share
Textual	Financial markets	Financial-market news topics	Monthly	Topic attention share
Textual	Government	Government and regulation topics	Monthly	Topic attention share
Textual	Industry	Industry and firm-level topics	Monthly	Topic attention share
Textual	International affairs	International news topics	Monthly	Topic attention share
Textual	Labor and income	Labor, income and tax topics	Monthly	Topic attention share
Textual	Oil and mining	Oil, mining and commodity topics	Monthly	Topic attention share

The number of lags and the polynomial degrees for the different MIDAS models are chosen to match those used in [1]. Accordingly, the monthly macroeconomic predictors have 12 lags, while the quarterly macroeconomic predictors have 4 lags, and the financial predictors and textual predictors have 3 lags. For the Legendre polynomial specification, the macroeconomic variables are of degree 3, whereas both the financial and textual variables are of degree 2. The significance of the predictors is evaluated using the regularization and variable selection methods described in Chapter 3. The predictive performance of these models is then compared using the aSPA test, RMSE, MSFE, and MAE. The RMSE, MSFE, and MAE can be expressed mathematically as

$$\text{MSFE} = \frac{1}{T}\sum_{t=1}^{T}(e_t)^2, \quad \text{RMSE} = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(e_t)^2}, \quad \text{MAE} = \frac{1}{T}\sum_{t=1}^{T}|e_t|.$$

In this thesis, we focus on constructing different models to forecast U.S. GDP growth. The data set consists of 289 quarterly observations of the adjusted value of the goods and services produced by labor and property located in the United States from March 1, 1947, to June 1, 2019. The annualized quarterly growth rate is computed as follows.

$$y_t = 100\left[\left(\frac{\text{RGDP}_t}{\text{RGDP}_{t-1}}\right)^4 - 1\right],$$

where y_t is the annualized quarterly real GDP growth rate in period t, RGDP_t is the real GDP in quarter t, and RGDP_{t-1} is the real GDP in the previous quarter.

The forecasting procedure is carried out in a rolling pseudo-out-of-sample framework. The first forecast is generated for 2002Q1 based on an initial estimation window of 60 quarters. The models are subsequently re-estimated in a rolling window, where the estimation sample is advanced by one quarter before each new forecast is produced. Forecast performance is evaluated over the out-of-sample period from 2002Q1 to 2017Q2, with the endpoint determined by the availability of the textual news variables. The ARDL-(U)-MIDAS models used in this application can be expressed as either the unrestricted MIDAS model or the MIDAS model with a Legendre polynomial.

$$y_t = \beta_0 + \sum_{p=1}^{P} \rho_p y_{t-p} + \sum_{r=1}^{R} \sum_{k=0}^{K_r} \beta_{r,k} x_{r,t,k} + \epsilon_t.$$

$$y_t = \beta_0 + \sum_{p=1}^{P} \rho_p y_{t-p} + \sum_{r=1}^{R} \sum_{\ell=0}^{L_r} \theta_{r,\ell} \left(\sum_{k=0}^{K_r} \phi_\ell(u_k) x_{r,t,k} \right) + \epsilon_t,$$

An illustration of quarterly U.S. real GDP growth from 1951Q1 to 2001Q4 is shown in Figure 4.1.

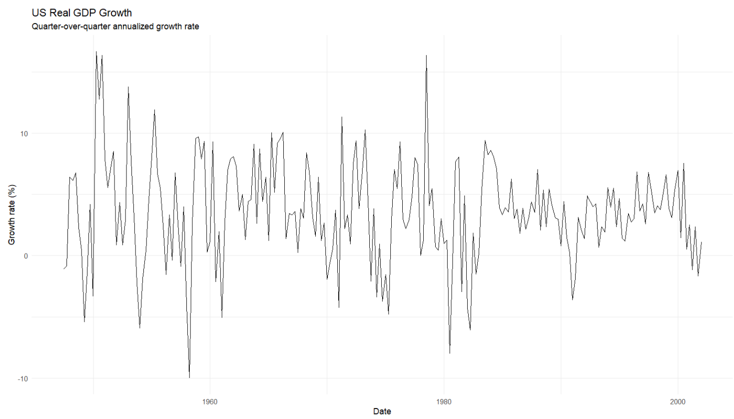


Figure 4.1: Quarterly U.S. GDP growth rate from 1951 to 2001.

Figure 4.1 shows the quarterly growth rate from the first quarter of 1951 to the last quarter of 2001. The figure shows significant volatility, particularly at the start and in the middle of the period. It stabilizes more at the end of the period. The volatility in the early part of the sample is partly attributable to recessions.

The volatility from 1960 to the early 1980s is due to the Great Inflation period, during which inflation rose from 1% in 1964 to more than 14% in 1980. This period also contains the 1973-1974 oil shock, during which members of OAPEC imposed an oil embargo on the

U.S., contributing to a sharp increase in oil prices. Toward the end of the sample, U.S. GDP growth is more stable than in earlier decades, with a dip in 2001. This is due to the dot-com bubble burst, during which tech stock prices fell sharply and firms cut back on investment. Furthermore, the 9/11 attacks also occurred that year.

The following sections present the out-of-sample forecasting results for 2002Q1–2017Q2. However, we note that textual variable data first became available around 1984. Hence, we set the initial estimation window from the first quarter of 1987 to the last quarter of 2001 for the initial forecast. We also note a publication delay of the macroeconomic variables, which are recorded monthly. This means that the macro variables will always be one month older than the other variables, especially the textual and financial variables. The empirical results are presented from the earliest forecast to the latest forecast within the quarter. We therefore begin with the 2-month horizon, where the information set contains the least recent data, before moving to the 1-month and end-of-quarter horizons.

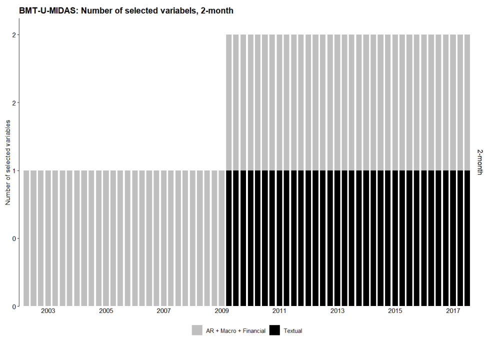
4.1 2-month horizon

This section evaluates the models at the 2-month horizon, where the information set is the least recent. This horizon is therefore the most challenging forecasting case and provides a benchmark for assessing how the methods perform when only early-quarter information is available. The analysis begins by comparing the number of selected predictors across methods. Figure 4.2 shows the number of variables selected by BMT, OCMT, LASSO, Elastic Net, and sparse-group LASSO at the 2-month horizon. The figure shows that the methods lead to very different selections. BMT is clearly the most restrictive method, as it selects only a small number of variables throughout most of the out-of-sample period.

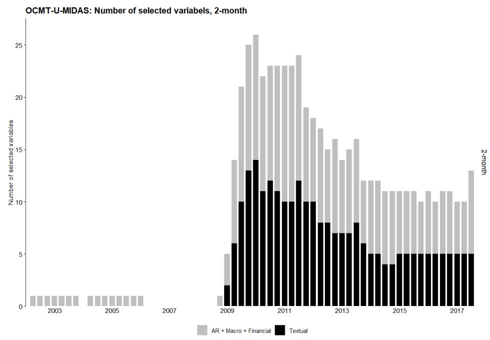
OCMT shows the opposite. It selects substantially more variables, especially after the mid-2000s. However, before 2009, it selected between zero and five variables each quarter.

LASSO displays a more irregular selection pattern. In some periods, it selects only a few variables, while in others the number of selected predictors increases. This suggests that the LASSO model complexity is quite sensitive to the time period. Elastic Net selects more variables than LASSO in most periods, consistent with its tendency to retain correlated predictors. Sparse-group LASSO selects the largest number of variables overall and therefore produces the most extensive predictor set among the methods considered.

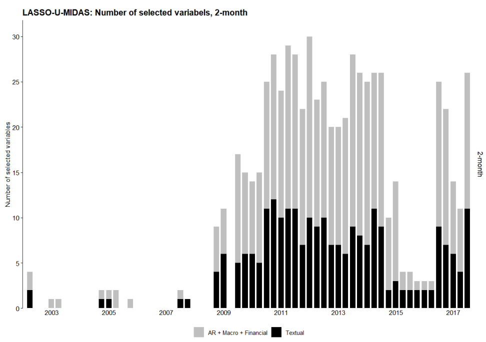
Overall, the figure shows that the methods differ not only in which variables they select, but also in the number of selected predictors. BMT is highly parsimonious, while OCMT is more selective than the penalized methods in some periods. In particular, around 2010, it selects many variables. LASSO is sparse but unstable over time, Elastic Net includes more variables, and sparse-group LASSO selects the most variables. Next, we examine which variables have been selected. A plot of the selected variables for the different methods is shown in Figure 4.3.



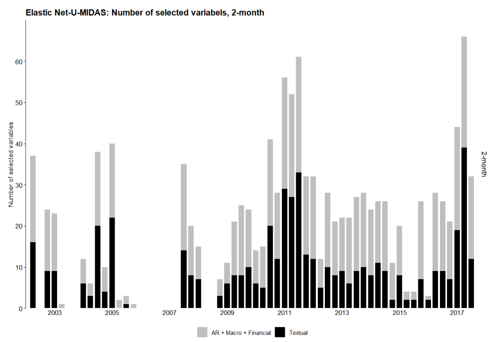
(a) BMT.



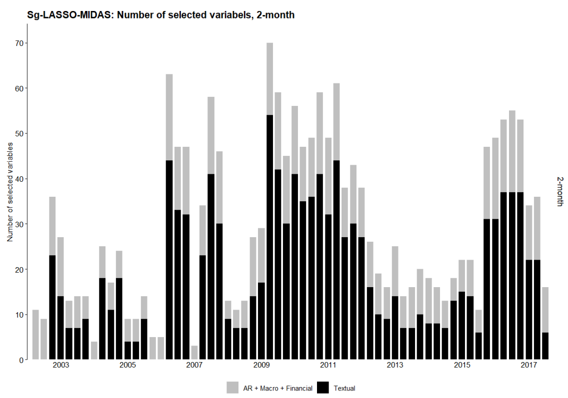
(b) OCMT.



(c) LASSO.

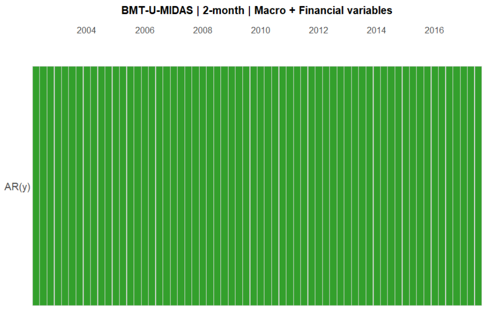


(d) Elastic Net.

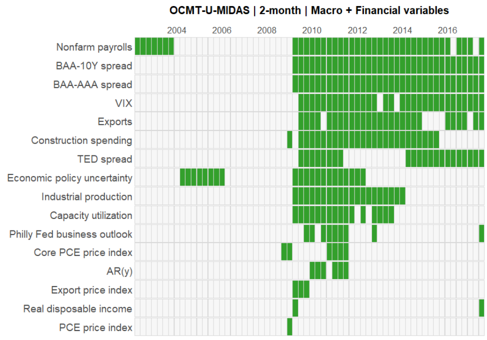


(e) Sparse-group LASSO.

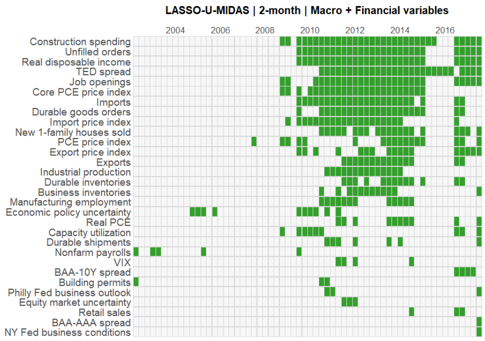
Figure 4.2: Number of selected variables by different methods at the 2-month horizon.



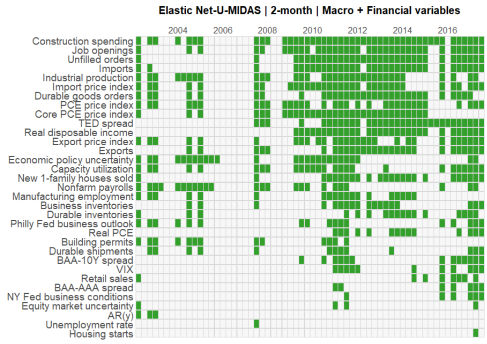
(a) BMT.



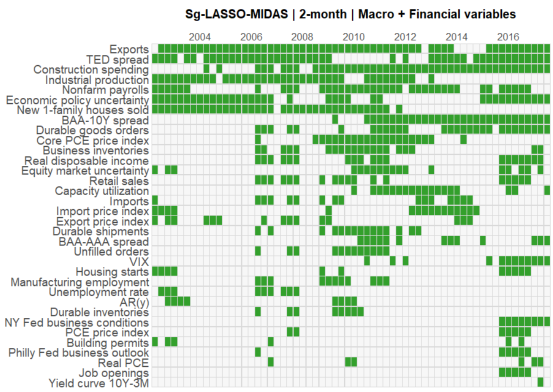
(b) OCMT.



(c) LASSO.



(d) Elastic Net.



(e) Sparse-group LASSO.

Figure 4.3: Macro-financial variable selection by different methods at the 2-month horizon.

Figure 4.3 shows that the different variable selection methods lead to substantially different predictors over time, as before. The green cells indicate periods in which a given macro and financial variable is selected, while the white cells indicate periods in which it is excluded.

BMT appears to select the fewest variables. It selects only one variable throughout the whole period. OCMT also produces a relatively sparse selection. Its selected variables are more concentrated in specific parts of the sample, especially later periods. This suggests that some predictors become relevant only during certain periods.

Compared with BMT and OCMT, LASSO selects a broader set of variables, although many of them are selected only temporarily. This points to a less stable selection pattern, where the importance of individual predictors changes frequently over time. Elastic Net selects an even larger number of variables, which is consistent with its tendency to retain groups of correlated predictors. This suggests that several macro and financial variables may contain overlapping information. Sparse-group LASSO provides a more structured selection pattern, selecting groups of variables, with some variables remaining selected for almost the entire period, as the exports and TED spread.

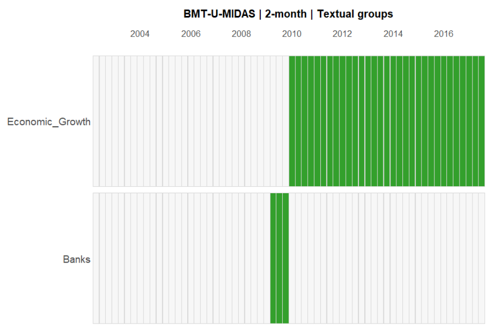
Overall, the figure indicates that predictor selection is both time-varying and method-dependent. Some methods identify only a few persistent predictors, while others retain a broader and more dynamic set of variables. Next, we examine which textual groups have been chosen by the different methods. Figure 4.4 illustrates the selection of textual groups across time for the 2-month forecasting horizon. The green cells again indicate that a given textual group is selected in a specific period, while the white cells indicate that it is excluded. The figure shows that the relevance of textual information is highly time-varying and differs substantially across the selection methods, as we saw with the macro variables.

BMT is the most restrictive method, selecting only very few textual groups, as with the macro variables. In particular, it identifies one group as relevant for a long period after around 2010, while another group is selected only briefly. OCMT is also relatively sparse, but it selects a broader set of textual groups than BMT. Its selected groups become especially persistent after the financial crisis period. This indicates that textual information may have become more useful for forecasting toward the end of the sample.

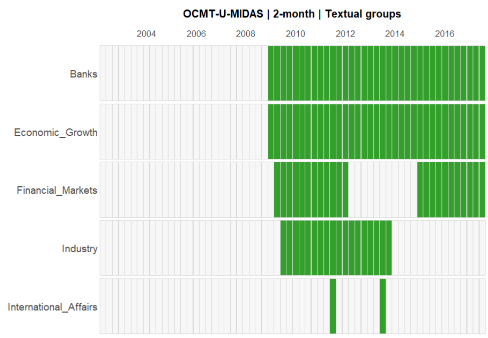
LASSO and Elastic Net select a larger number of textual groups. LASSO shows a more fragmented pattern, where several groups are selected intermittently, suggesting that the importance of textual predictors varies over time. Elastic Net selects groups more frequently and more persistently. This indicates that several textual categories may contain related information that is jointly useful for forecasting.

Sparse-group LASSO shows the broadest and most persistent selection pattern. Many textual groups are selected for long periods of the sample, especially from around 2009 onward. This indicates that the grouped structure in sparse-group LASSO helps capture broader patterns in the textual data, rather than selecting isolated textual predictors only. Groups such as economic growth, banking, financial markets, government, industry, and labor-related information appear to be selected repeatedly. This indicates that several textual groups contain relevant predictive variables.

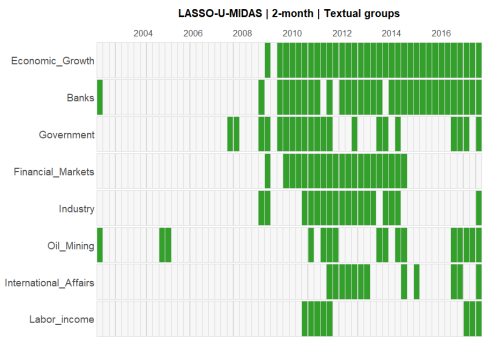
Overall, the figure suggests that textual groups become more important in the post-crisis period. BMT and OCMT produce relatively sparse selections, while Elastic Net and especially sparse-group LASSO retain more textual variables. The main conclusion is therefore that textual predictors appear to contain useful information about GDP growth.



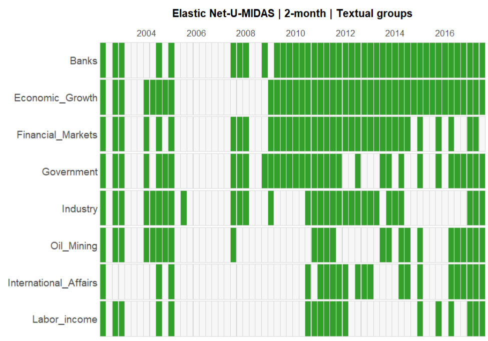
(a) BMT.



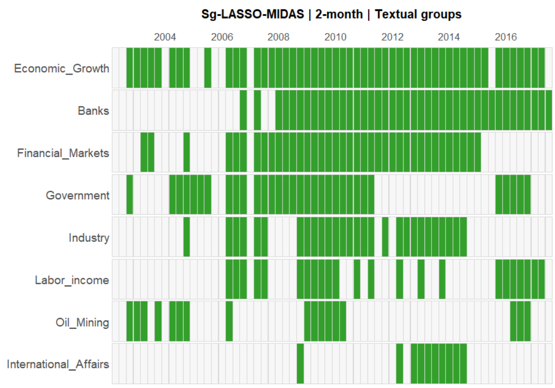
(b) OCMT.



(c) LASSO.



(d) Elastic Net.



(e) Sparse-group LASSO.

Figure 4.4: Textual group selection by different methods at the 2-month horizon.

The specific variables chosen by the methods for this horizon can be found in Figure B.2. To evaluate forecast performance, the RMSE, MSFE, and MAE are calculated for each method over the out-of-sample period. The resulting forecasts are plotted together with the actual GDP growth. This can be seen in Figure 4.5.

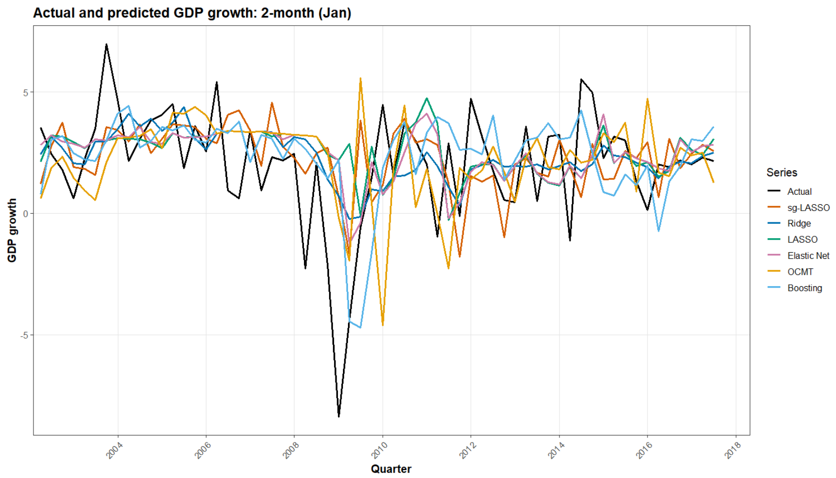


Figure 4.5: Predictions of quarterly GDP growth against actual quarterly GDP growth for the 2-month horizon.

Model	Horizon	RMSE	MSFE	MAE
Sg-LASSO-MIDAS	2-month	2.319	5.378	1.718
Ridge-U-MIDAS	2-month	**2.181**	**4.757**	**1.522**
Elastic Net-U-MIDAS	2-month	2.374	5.635	1.682
OCMT-U-MIDAS	2-month	2.713	7.359	1.924
BMT-U-MIDAS	2-month	2.417	5.840	1.743
LASSO-U-MIDAS	2-month	2.544	6.474	1.780

Table 4.2: Predictive performance of the different methods at the 2-month horizon.

Figure 4.5 shows the actual quarterly GDP growth rate with the predictions produced by the different methods at the 2-month-ahead horizon. The figure shows that the models generally follow the broad movements in GDP growth. However, the predictions are much smoother than the actual GDP growth. This is expected, since the models are estimated using a rolling window and rely on information available two months before the end of the quarter. Hence, the models are not expected to perfectly capture sudden changes in economic activity, especially when these changes occur very quickly.

The difference between the actual value and the forecasts is particularly clear during the financial crisis. During this period, actual GDP growth falls sharply, while all predictions decline by much less. This means that the models underestimate the size of the downfall. A similar pattern is also visible around the recovery after the crisis. Here, the actual GDP growth moves strongly, whereas the forecasts move more moderately. This suggests that the models are better at capturing normal fluctuations in GDP growth than extreme movements.

Another important feature of Figure 4.5 is that the forecasts from the different methods are relatively close to each other during stable periods. This indicates that when GDP growth is not changing dramatically, different selection and regularization methods yield similar predictions. The differences between the methods become more visible during periods of high volatility, where some forecasts react more strongly than others. This is also consistent with the variable selection plots above.

Methods such as OCMT select many variables and therefore offer greater flexibility, but this does not necessarily yield more accurate predictions. On the other hand, methods such as Ridge and sg-LASSO produce more regularized forecasts that appear more accurate over time.

Table 4.2 reports the RMSE, MSFE, and MAE for each model over the out-of-sample period. Across all three measures, Ridge-U-MIDAS performs best, with the lowest RMSE, MSFE, and MAE of 2.181, 4.757, and 1.522, respectively. This indicates that, at the 2-month-ahead horizon, the most accurate forecasts are obtained by shrinkage without exact variable selection.

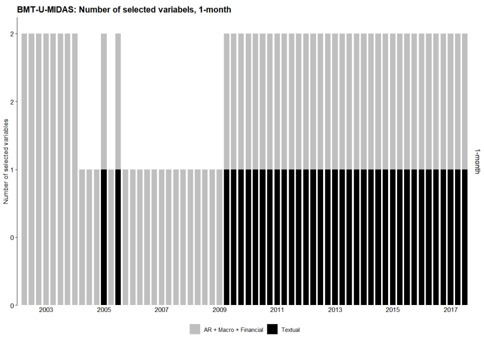
The second-best model by RMSE and MSFE is sg-LASSO-MIDAS, followed closely by Elastic Net-U-MIDAS. This suggests that methods that account for grouped predictors or correlated variables capture GDP growth more accurately. In contrast, LASSO-U-MIDAS performs worse, which may reflect its tendency to select only a subset of correlated predictors and omit useful variables.

BMT-U-MIDAS performs moderately well, outperforming LASSO and OCMT in terms of RMSE and MSFE. However, its low selection may exclude predictors relevant to forecasting GDP growth. OCMT-U-MIDAS has the weakest performance, with the highest RMSE, MSFE, and MAE. Although OCMT selected many variables in the selection plots, the results show that including more predictors does not necessarily improve forecasting accuracy.

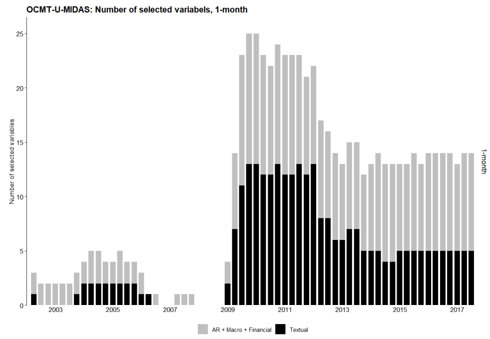
Overall, for the 2-month horizon, Ridge-U-MIDAS gives the best results. However, in the next section, the model uses information from one month earlier rather than from two months earlier.

4.2 1-month horizon

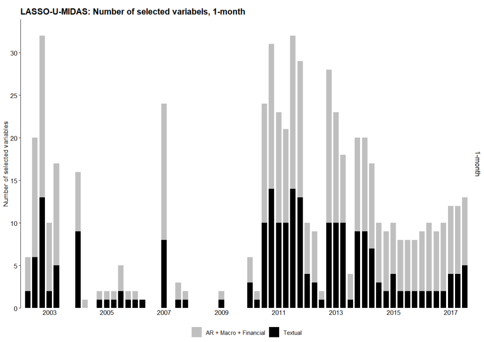
This section evaluates the models at the 1-month horizon. Compared with the 2-month horizon, the models now have access to more recent information within the quarter. The purpose is therefore to examine whether this additional information changes the selection patterns and improves forecast accuracy relative to the 2-month results. The number of selected variables for each method is shown in Figure 4.6.



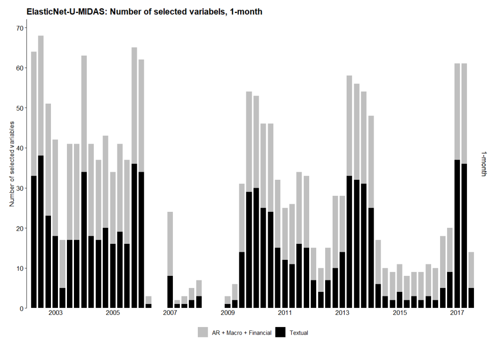
(a) BMT.



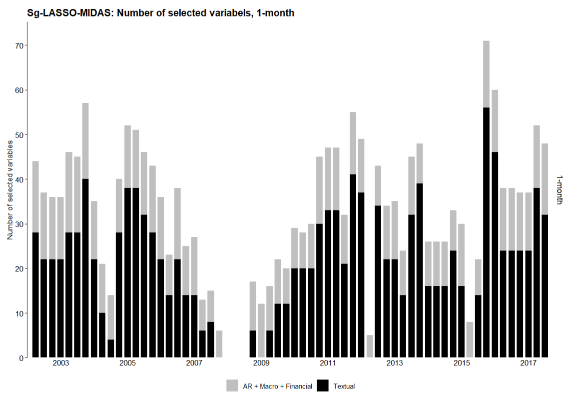
(b) OCMT.



(c) LASSO.



(d) Elastic Net.



(e) Sparse-group LASSO.

Figure 4.6: Number of selected variables by different methods at the 1-month horizon.

Figure 4.6 shows the number of selected variables at the 1-month horizon. The same pattern can be observed in this figure as in Figure 4.2. BMT remains the most parsimonious method, while OCMT selects a substantially larger number of variables, especially after the mid-2000s. Among the penalized methods, LASSO is still the most restrictive, whereas Elastic Net and sparse-group LASSO retain broader sets of predictors.

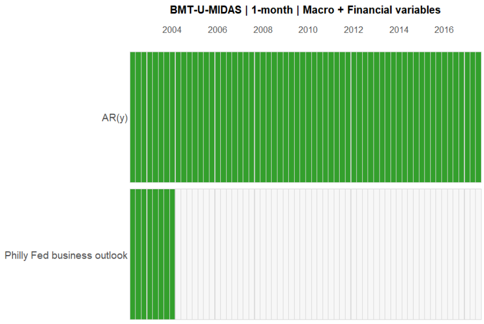
The main difference compared to the 2-month horizon is not a big change in the selection of variables, but rather a change in timing and persistence. Some variables become selected more frequently once more recent within-quarter information is available. This suggests that the horizon affects how consistent the selected predictors are, while the number of selected predictors mainly depends on the selection method

Next, we examine which macro and financial variables the different methods have chosen. Figure 4.7 shows the selected macro and financial variables at the 1-month horizon. The selection patterns are broadly similar to those in Figure 4.3. Variables related to labor-market conditions, production activity, construction, financial stress, uncertainty, and business surveys continue to appear across several methods.

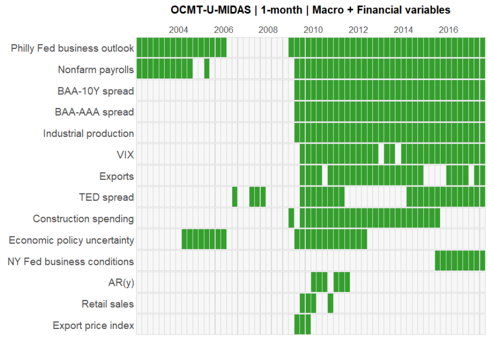
BMT continues to rely on only a few predictors, while OCMT selects a wider set of macro and financial variables in the model. LASSO selects a smaller and more time-varying subset, whereas Elastic Net and sparse-group LASSO retain broader groups of related variables. Thus, the 1-month horizon does not select a completely new set of relevant predictors. Instead, it strengthens the role of several variables that were already important at the 2-month horizon, including construction spending, industrial production, exports,job openings, credit spreads, the TED spread, the VIX, and business survey indicators.

Overall, the results indicate that the same broad information categories remain important when the forecast is made one month before the end of the quarter. However, their persistence depends on both the forecasting horizon and the variable-selection method. Next, we examine the textual groups instead. The selection of these can be seen in Figure 4.8. Figure 4.8 presents the selected textual groups at the 1-month horizon. The results are almost the same as in Figure 4.4. Economic growth and banking-related news remain among the most persistent textual groups, while financial markets, industry, oil and mining, labor and income, government, and international affairs are selected more often by the penalized methods.

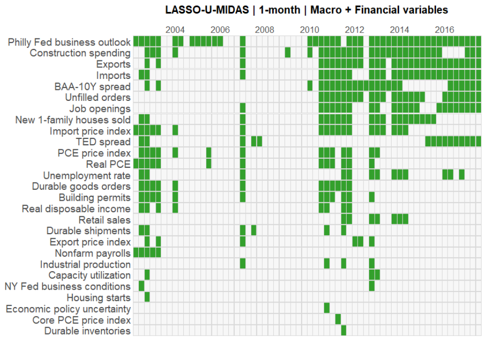
The main difference is that textual groups are selected more frequently and more persistently for some methods. This suggests that textual information is already useful early in the quarter. BMT and OCMT remain relatively selective, while Elastic Net and sparse-group LASSO assign a broader role to textual information.



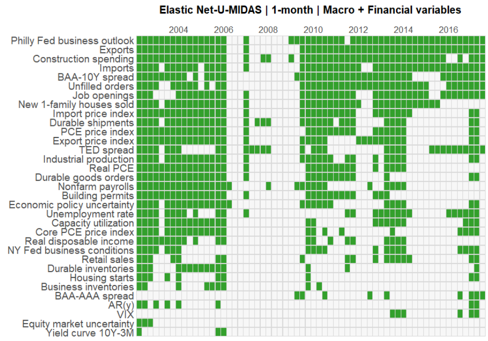
(a) BMT.



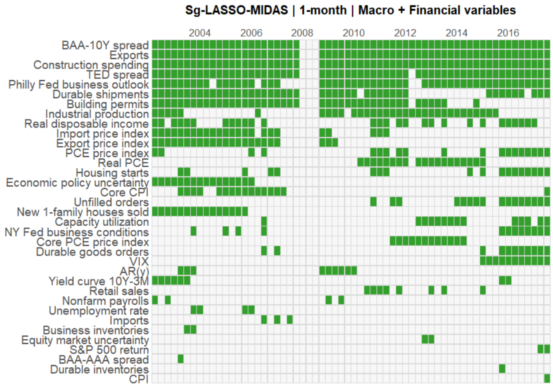
(b) OCMT.



(c) LASSO.

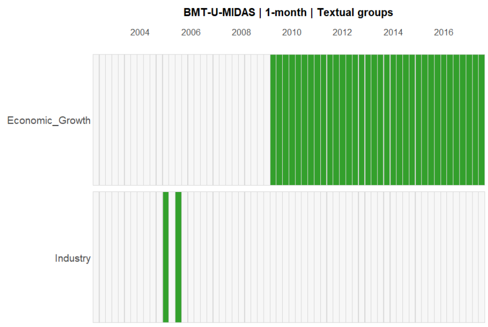


(d) Elastic Net.

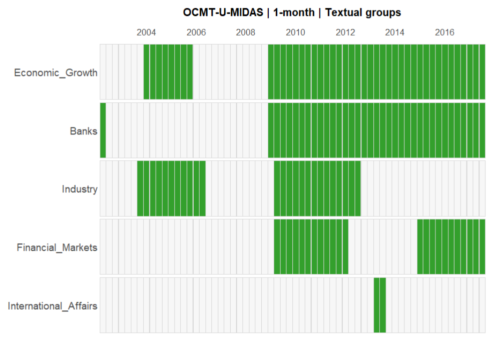


(e) Sparse-group LASSO.

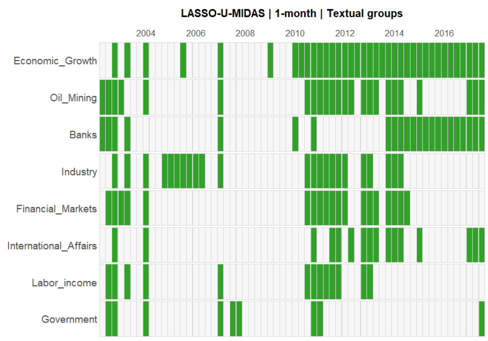
Figure 4.7: Macro-financial variable selection by different methods at the 1-month horizon.



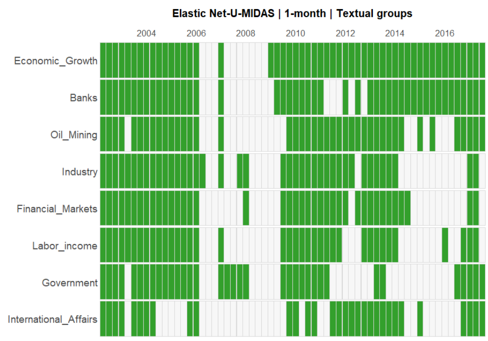
(a) BMT.



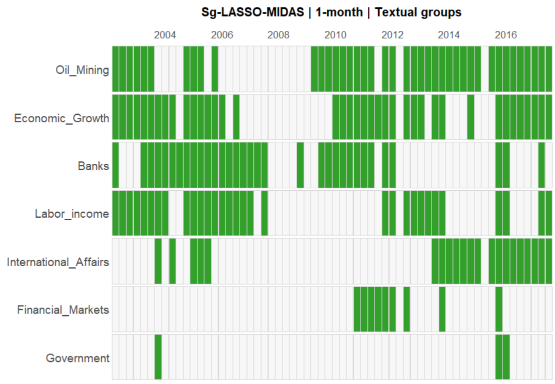
(b) OCMT.



(c) LASSO.



(d) Elastic Net.



(e) Sparse-group LASSO.

Figure 4.8: Textual group selection by different methods at the 1-month horizon.

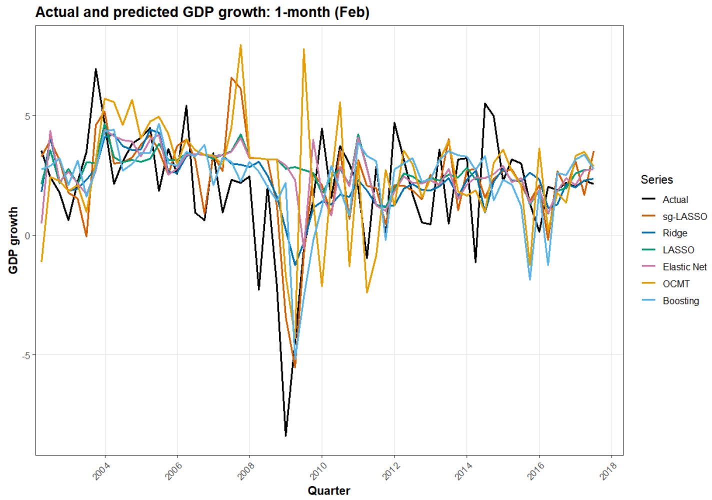


Figure 4.9: Predictions of quarterly GDP growth against actual quarterly GDP growth for the 1-month horizon.

Model	Horizon	RMSE	MSFE	MAE
Sg-LASSO-MIDAS	1-month	**2.022**	**4.088**	**1.549**
Ridge-U-MIDAS	1-month	2.192	4.805	1.565
BMT-U-MIDAS	1-month	2.394	5.730	1.729
LASSO-U-MIDAS	1-month	2.545	6.476	1.742
Elastic Net-U-MIDAS	1-month	2.580	6.655	1.794
OCMT-U-MIDAS	1-month	2.826	7.987	2.122

Table 4.3: Predictive performance of the different methods at the 1-month horizon.

Figure 4.9 shows the actual quarterly GDP growth rate together with the forecasts from the different methods at the 1-month horizon. Compared with Figure 4.5, the models now use more recent information within the quarter. This is reflected in the forecasts, which generally follow the realized GDP growth rate more closely than at the earlier horizon. However, the forecasts are still smoother than the actual series. This is especially evident during the financial crisis, where the decline in GDP growth is much sharper than predicted by the models. Hence, although the shorter horizon improves the information set, the models still struggle to capture sudden and extreme movements in economic activity.

The comparison with the 2-month horizon also shows that the differences between the methods become more visible when the forecast is made closer to the end of the quarter. At the 2-month horizon, Ridge-U-MIDAS performed best, suggesting that broad shrinkage was useful when the information set was still relatively limited. At the 1-month horizon, sg-LASSO-MIDAS becomes the best-performing model. This suggests that the grouped variable-selection structure becomes more useful once more recent macro-financial and textual information is available.

The forecasts from the different methods are relatively similar during stable periods, but they differ more when GDP growth becomes volatile. OCMT fluctuates more than the other forecasts. In contrast, Ridge-U-MIDAS and sg-LASSO-MIDAS produce smoother and more accurate forecasts. This indicates that regularization remains important at the 1-month horizon, even though the models have access to more recent information than in the 2-month case.

Table 4.3 confirms this pattern. Sg-LASSO-MIDAS performs best at the 1-month horizon, with the lowest RMSE, MSFE, and MAE, equal to 2.022, 4.088, and 1.549, respectively. Ridge-U-MIDAS is the second-best model and performs very similarly in terms of MAE, but its RMSE and MSFE are higher. This indicates that sg-LASSO-MIDAS produces fewer large forecast errors at the 1-month horizon compared to Table 4.2.

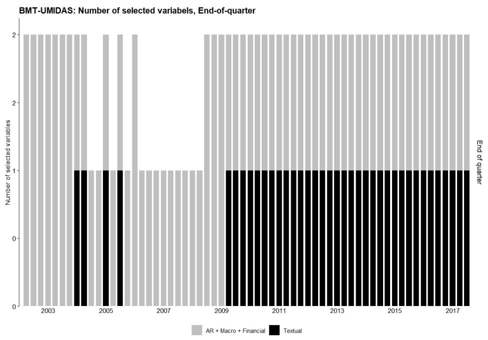
BMT-U-MIDAS performs better than LASSO-U-MIDAS, Elastic Net-U-MIDAS, and OCMT-U-MIDAS, but it does not match the performance of sg-LASSO-MIDAS or Ridge-U-MIDAS. OCMT has the weakest performance across all three measures, with an RMSE of 2.826, an MSFE of 7.987, and an MAE of 2.122. This is consistent with Table 4.2, where OCMT also performed relatively poorly despite selecting many variables. Thus, selecting a larger number of predictors does not necessarily improve out-of-sample forecasting accuracy.

Overall, the 1-month results show that forecast accuracy improves for BMT-U-MIDAS and sg-LASSO when the information set becomes more recent. Compared with the results in Table 4.2, the best-performing method changes from Ridge-U-MIDAS to sg-LASSO-MIDAS. This suggests that pure shrinkage is particularly useful for the earlier horizon, whereas sparse-group regularization is more effective at the 1-month horizon. Next, we move on to the final horizon, specifically the end-of-quarter horizon, which uses the most recent information.

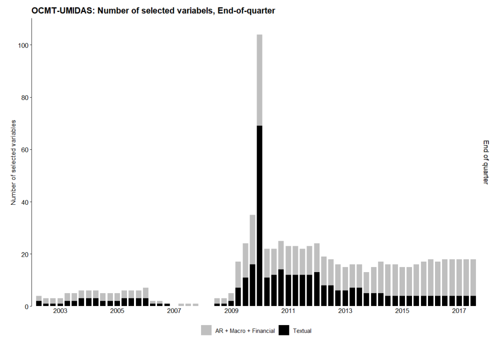
4.3 End-of-quarter horizon

In this section, the end-of-quarter horizon is considered. This horizon uses the most recent information available within the quarter. Compared with the 1-month and 2-month horizons, the models should have access to more recent data. As before, we begin by examining the number of variables selected by each method over the out-of-sample period. The number of selected variables for each method is shown in Figure 4.10.

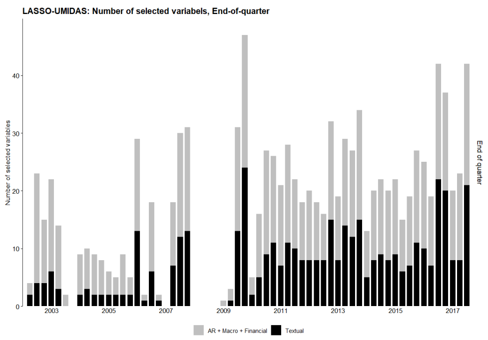
Figure 4.10 shows the number of selected variables at the end-of-quarter horizon. Although this horizon uses the most recent information within the quarter, the number of selections by the different methods remains similar to the earlier horizons. BMT continues to select the fewest variables. OCMT selects more variables but varies over time, whereas Elastic Net and sparse-group LASSO generally retain the broadest sets of predictors.



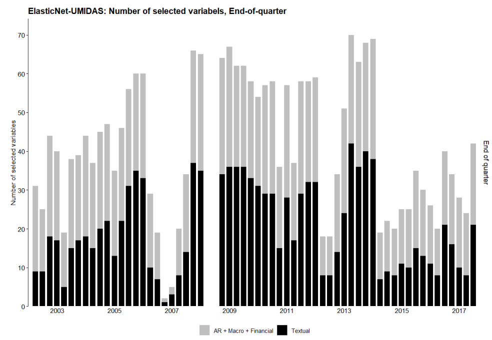
(a) BMT.



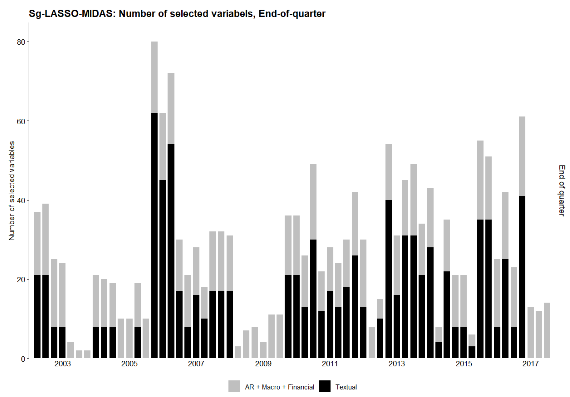
(b) OCMT.



(c) LASSO.



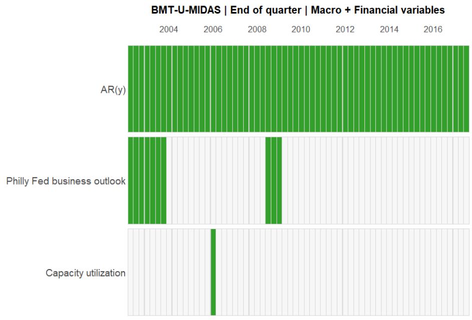
(d) Elastic Net.



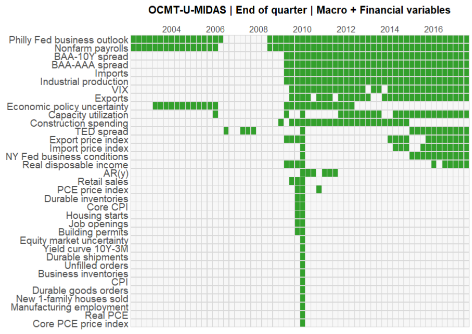
(e) Sparse-group LASSO.

Figure 4.10: Number of selected variables by different methods at the end-of-quarter horizon.

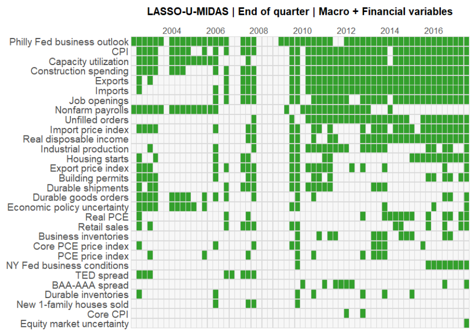
This indicates that the amount of within-quarter information does not fundamentally change the variable selection. Instead, variable selection primarily depends on the method used. The end-of-quarter horizon shows the same pattern as in Figure 4.6 and Figure 4.2. BMT selects very few predictors, OCMT selects more, LASSO remains fairly selective, and Elastic Net and sparse-group LASSO select broader sets of predictors. Next, we examine which macro variables have been selected over the period.



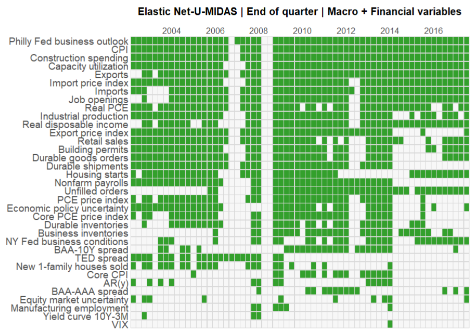
(a) BMT.



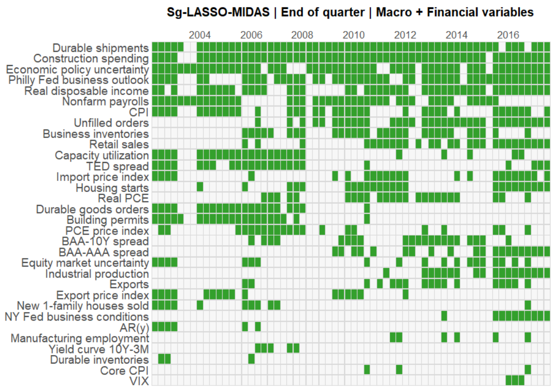
(b) OCMT.



(c) LASSO.



(d) Elastic Net.



(e) Sparse-group LASSO.

Figure 4.11: Macro-financial variable selection by different methods at the end-of-quarter horizon.

Figure 4.11 shows the selected macro-financial variables at the end-of-quarter horizon. The same broad variables remain important across all three horizons. These include the autoregressive component, construction spending, industrial production, exports, imports, job openings, credit spreads, the TED spread, the VIX, nonfarm payrolls, and business survey indicators.

The end-of-quarter horizon does not show an entirely different set of predictors compared to Figure 4.7 and Figure 4.3. Rather, some of the same variables become more or less persistent as the information set becomes more recent. BMT still selects only a few predictors, while OCMT continues to rely on a larger set of macro and financial variables. LASSO, Elastic Net, and sparse-group LASSO show a more balanced pattern, with Elastic Net and sparse-group LASSO selecting the widest number of groups.

Overall, the figure suggests that the selection of macro and financial variables is the same across horizons, but the exact variables selected depend most on the method used. Next, we look at the selected textual groups for each method. The specific variables chosen by the methods can be found in Figure B.4. Figure 4.12 reports the selected textual groups at the end-of-quarter horizon. Textual information remains relevant, but the pattern is somewhat more concentrated around macroeconomic and financial themes. Economic growth and banks are again among the most persistent groups, while financial markets, industry, oil and mining, labor and income, government, and international affairs are mainly selected by the penalized methods.

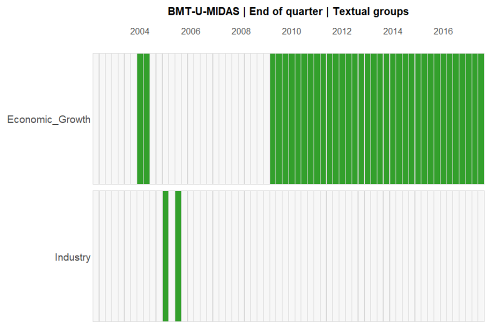
Compared with earlier horizons, the end-of-quarter results do not suggest that new textual groups are becoming dominant. Instead, the same groups keep appearing, but some last longer than others. This confirms that textual information is useful throughout the quarter. Next, we evaluate the forecasting performance of the different methods at the end-of-quarter horizon.

Figure 4.13 shows the actual quarterly GDP growth rate together with the forecasts from the different methods at the end-of-quarter horizon. Compared with the 2-month and 1-month horizons, this horizon uses the most recent information available within the quarter. The forecasts are therefore expected to follow realized GDP growth more closely. The figure shows that most methods capture the general movements in GDP growth, especially during stable periods. However, as in the previous horizons, the forecasts remain smoother than the actual series and continue to struggle to capture sudden, large changes in GDP growth.

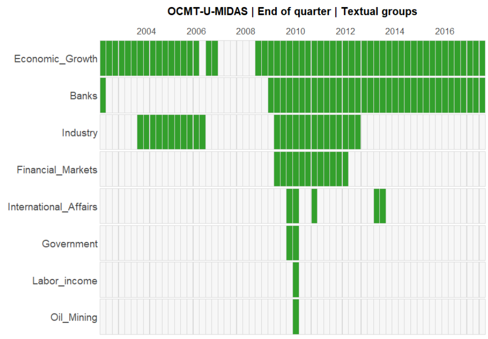
The largest forecast errors again occur around the financial crisis. This pattern was also visible in Figure 4.9 and Figure 4.5. Even when the models use information from the end of the quarter, they still react more moderately than the realized GDP series. The

Model	Horizon	RMSE	MSFE	MAE
Ridge-U-MIDAS	End-of-quarter	**2.026**	**4.105**	**1.499**
Sg-LASSO-MIDAS	End-of-quarter	2.037	4.147	1.528
BMT-U-MIDAS	End-of-quarter	2.079	4.321	1.563
Elastic Net-U-MIDAS	End-of-quarter	2.082	4.336	1.574
LASSO-U-MIDAS	End-of-quarter	2.562	6.561	1.773
OCMT-U-MIDAS	End-of-quarter	3.003	9.016	2.209

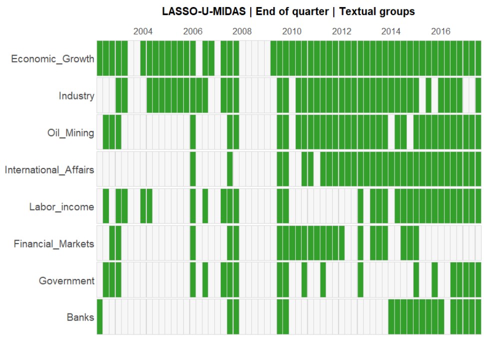
Table 4.4: Predictive performance of the different methods at the end-of-quarter horizon.



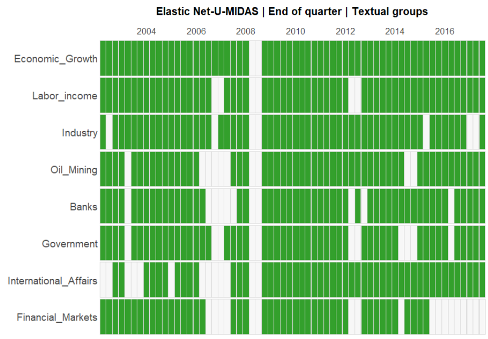
(a) BMT.



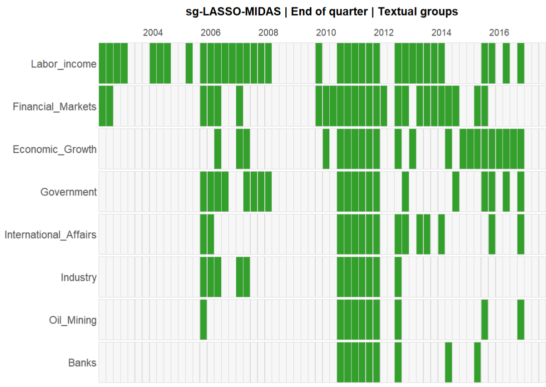
(b) OCMT.



(c) LASSO.



(d) Elastic Net.



(e) Sparse-group LASSO.

Figure 4.12: Textual group selection by different methods at the end-of-quarter horizon.

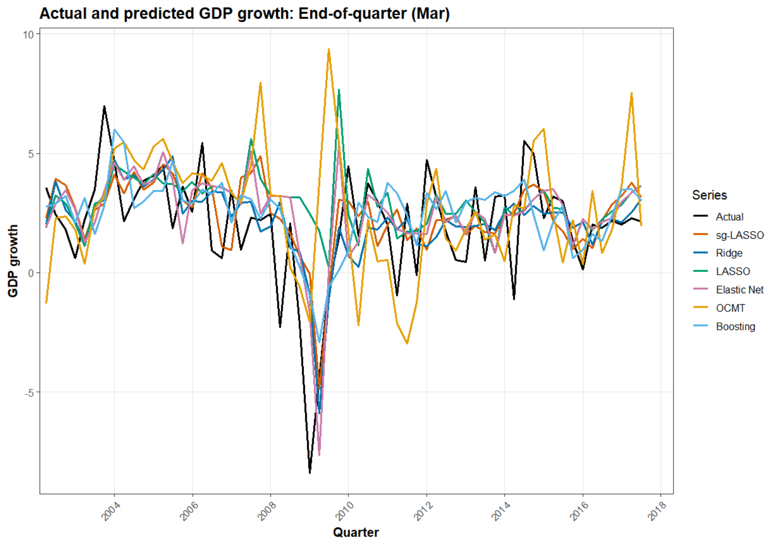


Figure 4.13: Predictions of quarterly GDP growth against actual quarterly GDP growth for the end-of-quarter horizon.

actual GDP growth falls sharply during the crisis, while the forecasts adjust more slowly. This suggests that the additional within-quarter information improves the forecasts during normal periods, but it does not capture the extreme macroeconomic shocks.

Compared with the results in Table 4.3, the end-of-quarter forecasts are not better for all methods. Ridge-U-MIDAS, BMT-U-MIDAS, and Elastic Net-U-MIDAS improve compared to Table 4.3, while sg-LASSO-MIDAS performs almost the same as before. In contrast, LASSO-U-MIDAS and OCMT-U-MIDAS perform worse at the end-of-quarter horizon than at the 1-month horizon. This indicates that more recent information does not necessarily improve forecasting performance.

The forecasts from Ridge-U-MIDAS, sg-LASSO-MIDAS, BMT-U-MIDAS, and Elastic Net-U-MIDAS are relatively close to each other throughout most of the sample. This is also reflected in Table 4.4, where these four methods have similar forecast errors. In contrast, OCMT-U-MIDAS is more volatile and deviates more from the actual series in several periods. This is consistent with the earlier horizons, where OCMT also performed poorly. Thus, the results again show that selecting a larger number of variables does not necessarily improve out-of-sample forecast accuracy.

Table 4.4 shows that Ridge-U-MIDAS is the best-performing model at the end-of-quarter horizon. It has the lowest RMSE, MSFE, and MAE, equal to 2.026, 4.105, and 1.499, respectively. Sg-LASSO-MIDAS performs almost equally well, with an RMSE of 2.037, an MSFE of 4.147, and an MAE of 1.528. Compared with the 1-month horizon, where sg-LASSO-MIDAS performed best, Ridge-U-MIDAS is again the best performing method. This is similar to the 2-month horizon, where Ridge-U-MIDAS also performed best. Hence, Ridge-U-MIDAS is the strongest method at both the earliest and the latest horizons, while sg-LASSO-MIDAS performs best at the 1-month horizon.

BMT-U-MIDAS and Elastic Net-U-MIDAS also perform relatively well at the end-of-quarter horizon. BMT-U-MIDAS obtains an RMSE of 2.079, an MSFE of 4.321, and an MAE of 1.563, while Elastic Net-U-MIDAS obtains an RMSE of 2.082, an MSFE of 4.336, and an MAE of 1.574. These values are close to the two best methods, suggesting that both BMT and Elastic Net make good use of the end-of-quarter information. In particular, BMT improves relative to the 1-month horizon.

LASSO-U-MIDAS and OCMT-U-MIDAS perform considerably worse. LASSO-U-MIDAS has an RMSE of 2.562, an MSFE of 6.561, and an MAE of 1.773. OCMT-U-MIDAS has the weakest performance across all three measures, with an RMSE of 3.003, an MSFE of 9.016, and an MAE of 2.209. Compared with the results in Table 4.3 and Table 4.2, OCMT remains the least accurate method.

Overall, the end-of-quarter results show that using the most recent within-quarter information improves forecasting performance relative to the 2-month horizon for several methods. However, the improvement relative to the 1-month horizon is small and not visible across all models. Ridge-U-MIDAS performs best at the end-of-quarter horizon, while sg-LASSO-MIDAS performs best at the 1-month horizon. Across all three horizons, the strongest forecasting performance is obtained by Ridge-U-MIDAS and sg-LASSO-MIDAS. This suggests that both shrinkage and structured regularization are important for producing accurate GDP forecasts. In contrast, OCMT-U-MIDAS is the weakest method across all horizons, indicating that a larger selected model does not necessarily yield better forecasting performance.

Next, we use the aSPA test to examine whether the models differ significantly in predictive accuracy, both jointly across all horizons and separately for each horizon.

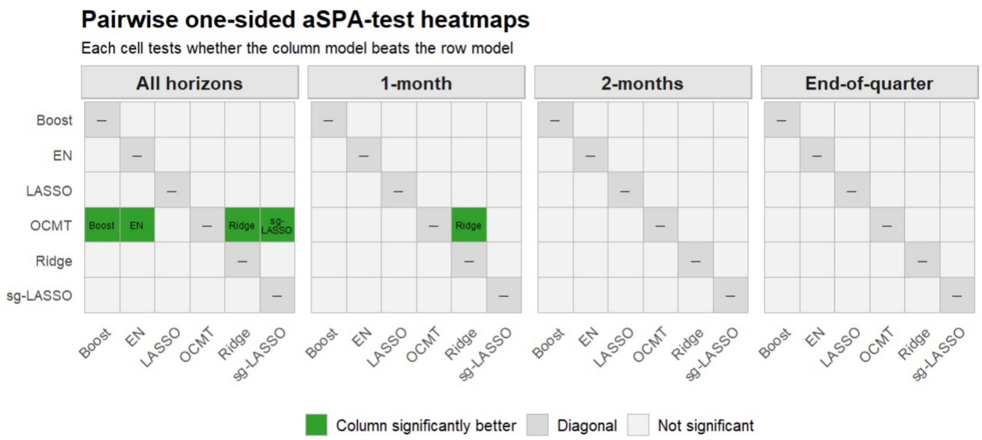


Figure 4.14: Pairwise one-sided aSPA-test heatmaps across horizons.

Figure 4.14 shows the results for the pairwise one-sided aSPA test. Each cell tests whether the column model has superior predictive ability relative to the row model. Across all horizons, several models perform significantly better than OCMT-U-MIDAS. These include BMT-U-MIDAS, Elastic Net-U-MIDAS, Ridge-U-MIDAS, and sg-LASSO-MIDAS. This confirms the earlier finding that OCMT-U-MIDAS is the weakest forecasting method overall.

At the individual horizons, however, there are fewer significant differences between the models. The only clear horizon-specific rejection appears at the 1-month horizon, where Ridge-U-MIDAS significantly outperforms OCMT-U-MIDAS. For the 2-month and end-of-quarter horizons, the aSPA test does not indicate statistically significant pairwise differences at the chosen significance level. Although Ridge-U-MIDAS and sg-LASSO-MIDAS have the lowest forecast error, Figure 4.14 shows that their advantage is not always statistically significant when the horizons are tested separately.

5 | Discussion

This chapter discusses the main choices, limitations, and possible extensions of the empirical analysis. The purpose of the thesis has been to compare different regularization and variable selection methods for MIDAS models. An important methodological choice is the use of MIDAS models with Legendre polynomial transformations. The Legendre polynomial provides a flexible and parsimonious way of summarising the lag structure of the high-frequency predictors. However, the use of Legendre polynomials is only one possible choice. Other polynomial choices could have been considered, such as Almon polynomials, Beta polynomials, or other orthogonal polynomial bases. Different choices may imply different lag shapes and may therefore affect both variable selection and predictive accuracy. A possible extension would therefore be to compare the results across several polynomials.

Another possible extension is to compare the MIDAS-based machine learning models with factor-augmented models. Dynamic factor models are commonly used in macroeconomic forecasting because they summarize a large number of variables using a smaller number of factors. In this thesis, the focus has instead been on regularization and variable selection within MIDAS models. A factor-augmented MIDAS approach could combine these two ideas by extracting common factors from the macroeconomic, financial, or textual variables and then using these factors in a MIDAS regression. Such an extension would also make it possible to compare whether variable selection or factor-based dimension reduction has better predictive accuracy

The data choice is another limitation. The empirical application includes macroeconomic, financial, and textual news variables, which makes it possible to compare the contribution of different types of information. However, the predictor set could be expanded further. Additional financial indicators, survey measures, payment data, labor market indicators, or alternative textual measures could be included.

A further limitation concerns the out-of-sample period. The forecasting horizon is from 2002Q1 to 2017Q2, which means that the financial crisis is included in the evaluation period. The financial crisis is an unusual period with large movements in GDP growth, and the forecast plots show that the models have difficulty capturing the sharp downturn and the following recovery. This suggests that the models are better at forecasting normal fluctuations than extreme macroeconomic shocks. A useful robustness check would therefore be to repeat the empirical analysis without the financial crisis period.

The frequency of the data also introduces an important practical issue. The macroeconomic variables are released with a delay, whereas the financial and textual variables are available more quickly. Consequently, the macroeconomic predictors are effectively older than the financial and textual predictors at each horizon. This means that the information set is not perfectly balanced across variable groups. However, it also means that part of the difference between macroeconomic, financial, and textual variables may be due to publication timing rather than significance.

Another practical choice is how variables are treated when they are not yet available in the sample. In the implementation, variables that have not yet been introduced at the start of the training set are initialized to zero. This choice may affect the early part of the sample, since zeros may be interpreted by the models as actual observations rather than missing information. An alternative approach would be to allow the predictor set to change over time or to use missing-data methods that distinguish between unavailable observations and true zero values.

6 | Conclusion

The purpose of this thesis has been to compare regularization and variable-selection methods for high-dimensional MIDAS models used to forecast quarterly U.S. real GDP growth. The analysis combines macroeconomic, financial, and textual news variables observed at different frequencies. Since GDP is only observed quarterly, while many relevant predictors are observed at higher frequencies, MIDAS models provide a useful framework for incorporating timely information without first aggregating all variables to the same frequency.

The empirical application in Chapter 4 was carried out using a rolling pseudo-out-of-sample forecasting framework over the period 2002Q1 to 2017Q2. Forecasts were produced for three horizons, namely a 2-month horizon, a 1-month horizon, and an end-of-quarter horizon. The monthly macroeconomic predictors were included with 12 lags, the quarterly macroeconomic predictors with 4 lags, and the financial and textual variables with 3 lags. Legendre polynomial transformations were used to reduce the dimensionality of the lag structure, with degree 3 for the macroeconomic variables and degree 2 for the financial and textual variables. The methods considered were Ridge, LASSO, Elastic Net, sparse-group LASSO, OCMT, and BMT, and their forecasting performance was evaluated using RMSE, MSFE, MAE, and the aSPA test.

The results in Chapter 4 show that Ridge-U-MIDAS and sparse-group LASSO-MIDAS provide the strongest overall forecasting performance. At the 2-month horizon in Section 4.1, Ridge-U-MIDAS performs best across all three forecast measures, with an RMSE of 2.181, an MSFE of 4.757, and an MAE of 1.522. Sg-LASSO-MIDAS is the second-best model in terms of RMSE and MSFE, with values of 2.319 and 5.378, while Elastic Net-U-MIDAS obtains the second-lowest MAE of 1.682. BMT-U-MIDAS and LASSO-U-MIDAS perform less accurately, with RMSE values of 2.417 and 2.544, respectively. OCMT-U-MIDAS has the weakest performance at this horizon, with an RMSE of 2.713, an MSFE of 7.359, and an MAE of 1.924. This suggests that broad shrinkage is particularly useful when the information set is still relatively limited early in the quarter.

At the 1-month horizon in Section 4.2, sparse-group LASSO-MIDAS gives the best results, with an RMSE of 2.022, an MSFE of 4.088, and an MAE of 1.549. Ridge-U-MIDAS is the second-best model, with an RMSE of 2.192, an MSFE of 4.805, and an MAE of 1.565. BMT-U-MIDAS also performs competitively, with an RMSE of 2.394, an MSFE of 5.730, and an MAE of 1.729, and it performs better than LASSO-U-MIDAS, Elastic Net-U-MIDAS, and OCMT-U-MIDAS. LASSO-U-MIDAS and Elastic Net-U-MIDAS obtain RMSE values of 2.545 and 2.580, respectively, while OCMT-U-MIDAS again performs worst, with an RMSE of 2.826, an MSFE of 7.987, and an MAE of 2.122.

At the end-of-quarter horizon in Section 4.3, Ridge-U-MIDAS again performs best, with an RMSE of 2.026, an MSFE of 4.105, and an MAE of 1.499. Sparse-group LASSO-MIDAS performs almost equally well, with an RMSE of 2.037, an MSFE of 4.147, and an MAE of 1.528. BMT-U-MIDAS and Elastic Net-U-MIDAS also produce competitive forecasts, with RMSE values of 2.079 and 2.082, respectively. Their MSFE values are 4.321 and 4.336, while their MAE values are 1.563 and 1.574. In contrast, LASSO-U-MIDAS performs worse, with an RMSE of 2.562, an MSFE of 6.561, and an MAE of 1.773. OCMT-U-MIDAS is again the weakest model, with an RMSE of 3.003, an MSFE of 9.016, and an MAE of 2.209. Thus, Ridge-U-MIDAS performs best at both the earliest and latest horizons, while sparse-group LASSO-MIDAS performs best at the 1-month horizon.

The aSPA test results in Figure 4.14 indicate that Ridge-U-MIDAS, sparse-group LASSO-MIDAS, BMT-U-MIDAS, and Elastic Net-U-MIDAS significantly outperform OCMT-U-MIDAS on average across the three forecast horizons. At the 1-month horizon, the aSPA test also shows that Ridge-U-MIDAS significantly outperforms OCMT-U-MIDAS. In the remaining comparisons, there is no statistically significant difference in predictive performance.

The main conclusion is therefore that Ridge-U-MIDAS and sparse-group LASSO-MIDAS produce the lowest forecast errors overall, while OCMT-U-MIDAS has the weakest forecasting performance among the methods considered. The results suggest that the best forecasting accuracy is obtained either by applying broad shrinkage across many predictors or by exploiting the grouped lag structure in the MIDAS specification. At the same time, the variable-selection results show that selecting many predictors is not necessarily beneficial. In particular, OCMT-U-MIDAS often selects a relatively large set of variables but still performs poorly out of sample.

Bibliography

[1] Babii, A., Ghysels, E., and Striaukas, J. (2022). Machine learning time series regressions with an application to nowcasting. *Journal of Business & Economic Statistics*, 40(3):1094–1106.

[2] Bybee, L., Kelly, B., Manela, A., and Xiu, D. (2024). Business news and business cycles. *The Journal of Finance*, 79(5):3105–3147.

[3] Chudik, A., Kapetanios, G., and Pesaran, M. H. (2018). A one covariate at a time, multiple testing approach to variable selection in high-dimensional linear regression models. *Econometrica*, 86(4):1479–1512.

[4] Donoho, D. L. and Johnstone, I. M. (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81(3):425–455.

[5] Federal Reserve Bank of St. Louis (2026). FRED economic data. `https://fred.stlouisfed.org/`. Accessed: 20 May 2026.

[6] Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, 33(1):1–22.

[7] Ghysels, E., Santa-Clara, P., and Valkanov, R. (2004). The MIDAS touch: Mixed data sampling regression models. Working paper, University of North Carolina and UCLA.

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

[9] Kapetanios, G., Sarafidis, V., and Ventouri, A. (2026). Model selection in high-dimensional linear regression using boosting with multiple testing. arXiv preprint arXiv:2602.19705.

[10] Quaedvlieg, R. (2021). Multi-horizon forecast comparison. *Journal of Business & Economic Statistics*, 39(1):40–53.

Appendices

B | Appendices to chapter 4.

B.1 Data description

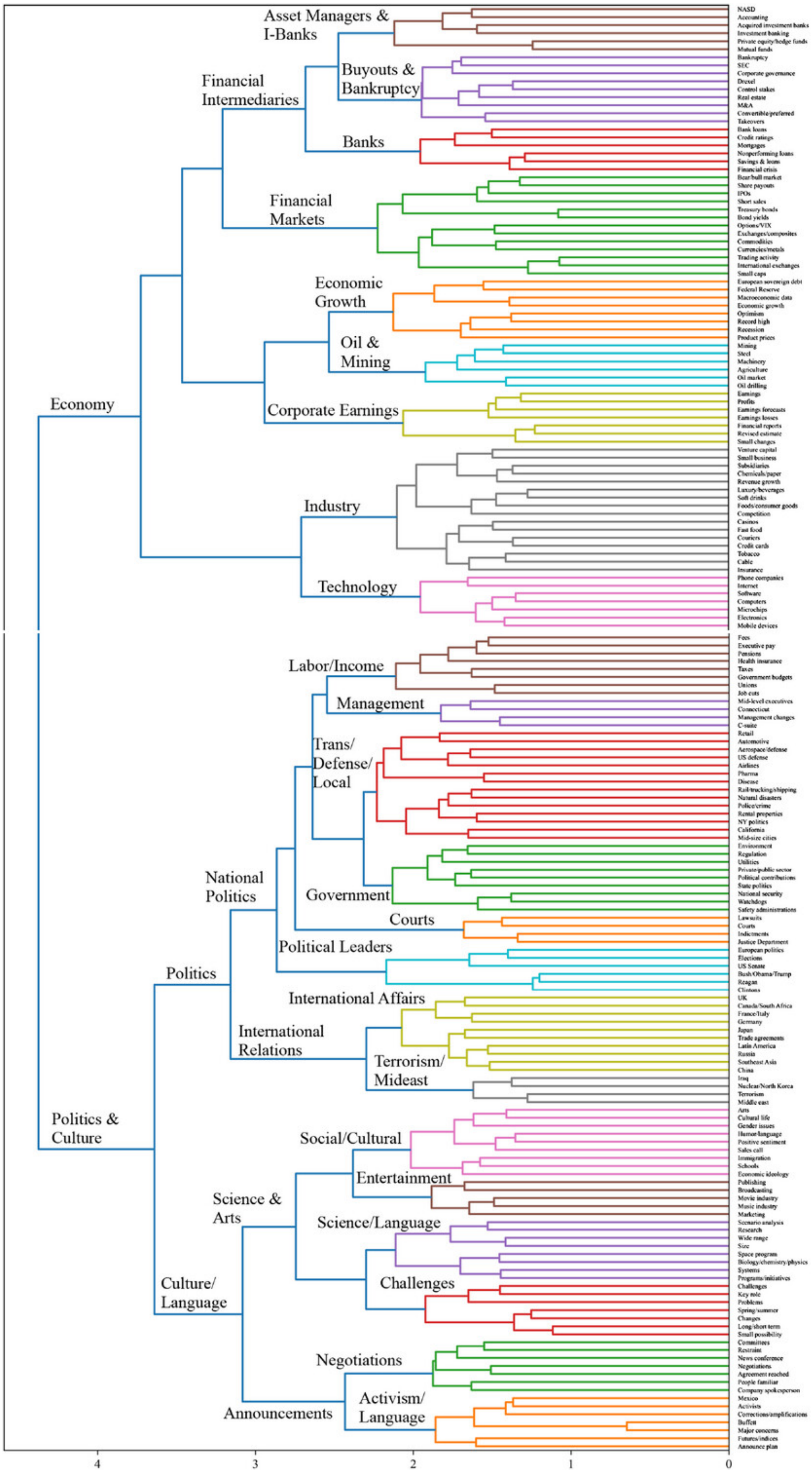
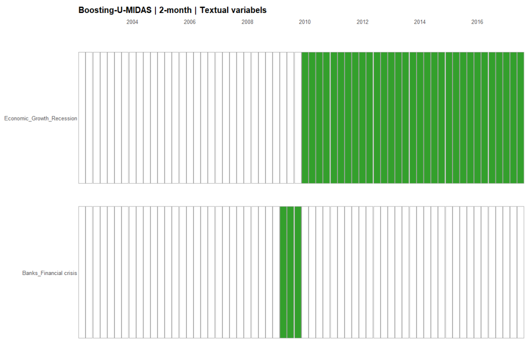
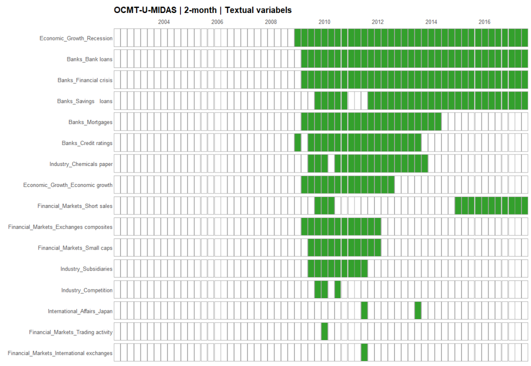


Figure B.1: Textual variables in the different groups

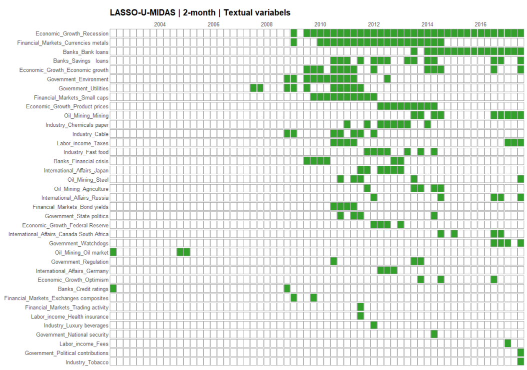
B.2 Selected textual variables



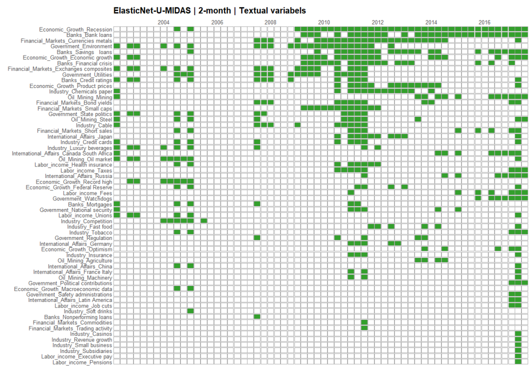
(a) Textual variable selection of BMT.



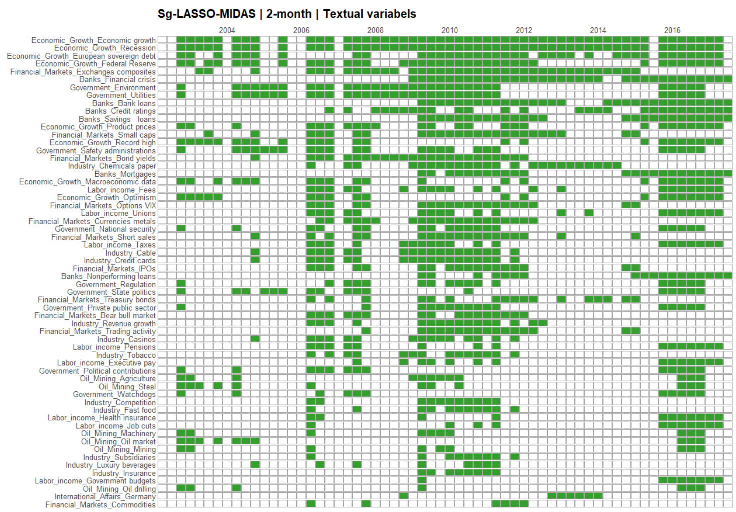
(b) Textual variable selection of OCMT.



(c) Textual variable selection of LASSO

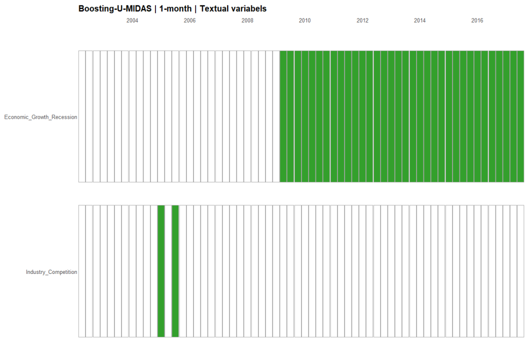


(d) Textual variable selection of Elastic net

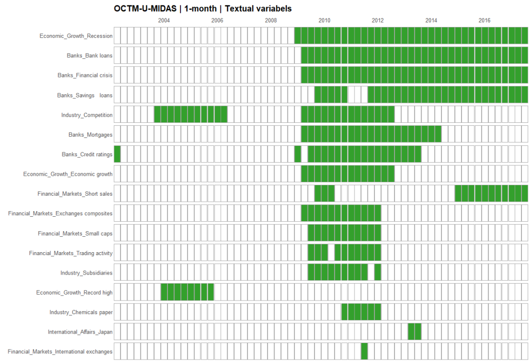


(e) Textual variable selection of Sparsegroup-LASSO

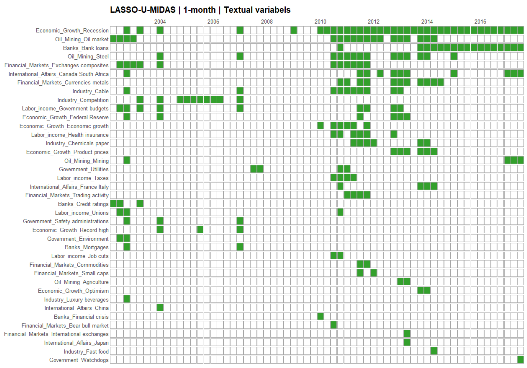
Figure B.2: Variable selection of different methods for the 2-month horizon



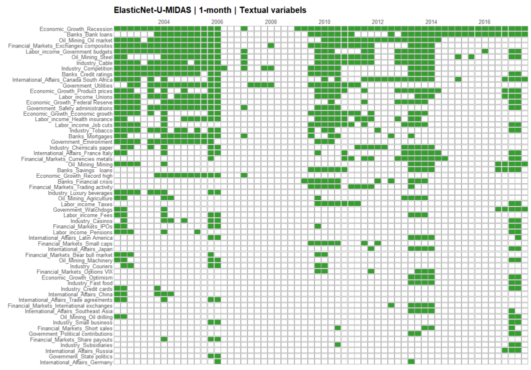
(a) Textual variable selection of BMT.



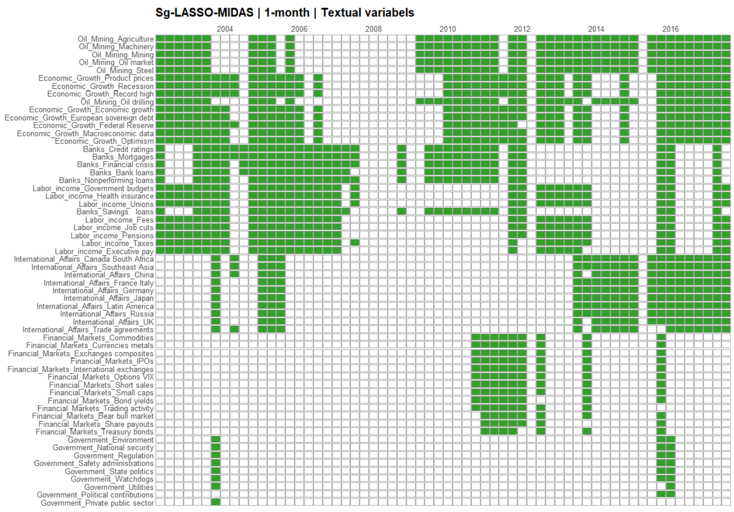
(b) Textual variable selection of OCMT.



(c) Textual variable selection of LASSO

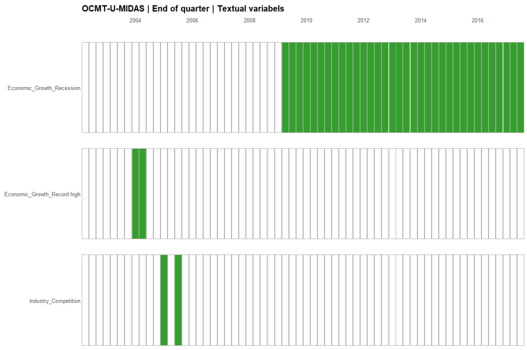


(d) Textual variable selection of Elastic net

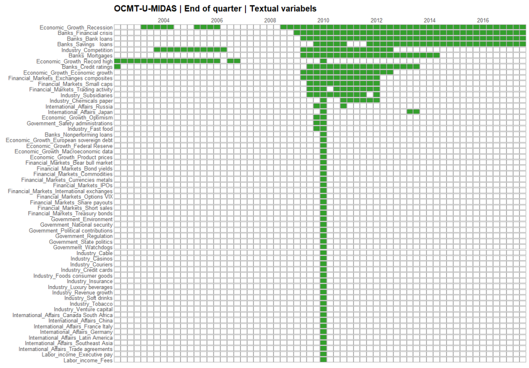


(e) Textual variable selection of Sparsegroup-LASSO

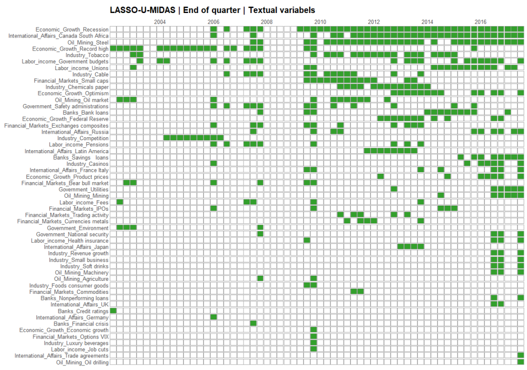
Figure B.3: Variable selection of different methods for the 1-month horizon



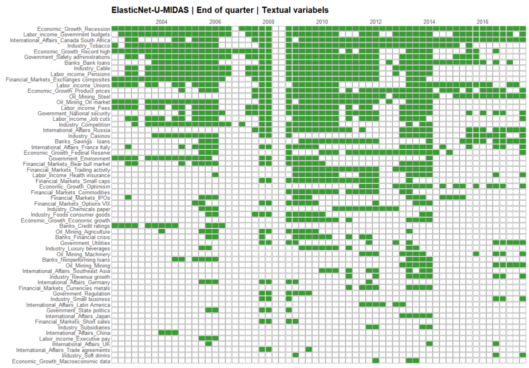
(a) Textual variable selection of BMT.



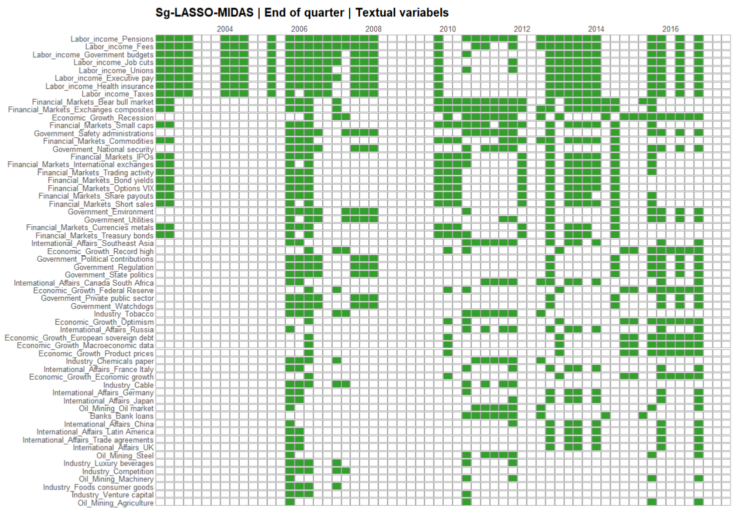
(b) Textual variable selection of OCMT.



(c) Textual variable selection of LASSO



(d) Textual variable selection of Elastic Net.



(e) Textual variable selection of Sparsegroup-LASSO

Figure B.4: Variable selection of different methods for the end-of-quarter horizon