Flexible Bayesian MIDAS: time-variation, group-shrinkage and sparsity*

David Kohns[†] Galina Potjagailo [‡]

Aalto University Bank of England

November 15, 2023

Abstract

We propose a mixed-frequency regression prediction approach that models a time-varying trend, stochastic volatility and fat-tails in the variable of interest. The coefficients of high-frequency indicators are regularized via a shrinkage prior that accounts for the grouping structure and within-group correlation among lags. A new sparsification algorithm on the posterior motivated by Bayesian decision theory derives inclusion probabilities over lag groups, thus making the results easy to communicate without imposing sparsity a priori. An empirical application on nowcasting UK GDP growth suggests that group-shrinkage in combination with the time-varying components substantially increases nowcasting performance by reading signals from an economically meaningful sub-set of indicators, whereas the time-varying components help by allowing the model to switch between indicators. Over the data release cycle, signals initially stem from survey data and then shift towards few "hard" real activity indicators. During the Covid-19 pandemic, the model performs relatively well since it shifts towards indicators for the service and housing sectors that capture the disruptions from economic lockdowns.

Keywords: Bayesian MIDAS regressions, Forecasting, Time-variation and fat tails, Grouped Horseshoe Prior, Decision Analysis.
JEL Codes: C11, C32, C44, C53, E37

*We are grateful for highly useful comments and suggestions from Arnab Bhattacharjee, Todd Clark, George Kapetanios, Gary Koop, Silvia Miranda-Agrippino, Andre Moreira, Tatevik Sekhposyan, and the participants of the 42nd International Symposium on Forecasting, the Bergamo BWES 2022 Workshop, the Conference on Real-Time Data Analysis, Methods, and Applications at the Cleveland Fed, the Orebro 5th Annual Workshop on Financial Econometrics, the 2023 SNDE Annual Symposium, and the RES 2023 Annual Conference. *Disclaimer:* The views expressed in this paper are solely those of the authors and do not necessarily represent those of the Bank of England or the European Central Bank.

†Corresponding author. Email: david.kohns94@googlemail.com

‡Email: galina.potjagailo@bankofengland.co.uk

1 Introduction

The large and unprecedented economic fluctuations caused by the Covid-19 pandemic have reemphasized the challenge for time series and prediction models to be flexible enough to extract heterogeneous signals and to rely on different sets of indicators over time, as well as to accommodate time variation and extreme observations (Lenza and Primiceri, 2022; Carriero et al., 2021; Antolin-Diaz et al., 2021). Timely survey indicators that were previously heavily relied upon for nowcasting were less able to capture the magnitude of the economic decline and the heterogeneous impact across economic sectors (Bank of England, 2020). Even as the world has emerged from the pandemic, this remains relevant since economists increasingly monitor heterogeneous types of new data sets, and need to incorporate large economic fluctuations into their models such as the recent surge in inflation rates in many economies.

We propose a Bayesian mixed-frequency regression model that aims to address these challenges. The T-SV-t BMIDAS models gradual trend shifts and potentially extreme variance shifts in the variable of interest via a time-varying trend, stochastic volatility and t-distributed errors that can assume fat tails. If left unmodeled, such variation in the variable of interest can blur nowcast or forecast signals. Information from higher-frequency indicators enters via a multivariate mixed-data sampling (MIDAS) regression that links the variable of interest to higher-frequency lags of each indicator via parametric functional constraints (Ghysels et al., 2007, 2020). The high-frequency lag coefficients are regularized via a flexible three-tiered group-wise shrinkage prior that jointly shrinks groups of lags of less informative indicators while accounting for the correlation among lags. Finally, we propose a sparsification algorithm for the posterior motivated by Bayesian decision theory which achieves ex-post variable selection with the aim of intuitive signal communication.

For group-shrinkage, we employ the GIGG (Group Inverse-Gamma Gamma) global-local prior, where two heavy-tailed processes shrink irrelevant signals towards zero while avoiding excessive shrinkage of relevant signals globally and locally (Polson and Scott, 2010). Proposed by Boss et al. (2021) for panel settings, the prior adds a group-wise shrinkage dimension that simultaneously shrinks *between* and *within groups* of covariates. We adopt the prior for the time-series context and set hyperparameters such that between-group shrinkage with high correlation within group is preferred. This regularizes variance inflation commonly observed in over-parameterized multivariate MIDAS models (Carriero et al., 2015) and accounts for the temporal lag correlation between higher-frequency observations of the same indicator (which we define as "groups").[1]

The prior shrinks in a continuous way without imposing exact zero coefficients, which makes it more flexible. However, a drawback is that communication of the signals that the model exploits is difficult without a clear-cut selection of indicators. To address this, we propose a new sparsification algorithm motivated by Bayesian decision theory that is applied to the posterior of the MIDAS coefficients. The algorithm selects *ex-post* those high-frequency lag groups that best summarize the predictions of the model, in the spirit of Hahn and Carvalho (2015), and sets the coefficients of others to exact zeros. Imposing this via the prior directly is known to be sensitive to the hyperparameters (Barbieri and Berger, 2004; O'Hara and Sillanpää, 2009) and to the correlation structure in the data (Barbieri et al., 2021). By asking which subsets of data reproduce the posterior predictions as closely as possible, the *ex-post* sparsification step allows to assign inclusion probabilities to coefficient groups and hence to enhance the interpretability and tractability of predictions.

In an empirical application, we nowcast quarter-on-quarter GDP growth in the United Kingdom using monthly macroeconomic indicators. We conduct nowcasts in a pseudo-real-time setting over the sample period 1999 to 2021, also examining a pre-pandemic sample. We find that the proposed model performs competitively against a range of alternatives for both sample periods. During the pandemic quarters, the model detects the initial economic trough earlier and nowcasts the economic recovery more precisely.

To understand the role of the main model features, we compare mean and density nowcast performance to models that shut some or all of the time-varying components down, as well as against a range of alternative priors on the MIDAS coefficients. We find that allowing for a time-varying trend and SV with t-distributed errors, substantially improves nowcast performance when combined with the group-shrinkage prior. Examining inclusion probability patterns, we show that the GIGG prior identifies a small set of indicators with high inclusion probability, while still reading signals from other indicators albeit with lower probability and higher uncertainty. Over the data release cycle, the model reads signals from survey indicators early on and then shifts towards signals from a few real activity indicators. Moreover, we find that the combination of the time-

[1] Intra-group collinearity is present in unrestricted MIDAS (Foroni et al., 2015), and is reinforced when polynomial restrictions are imposed to achieve parsimony and a meaningful lag profile (Ferrara et al., 2022). Generally, three-tiered group-shrinkage can be relevant in a broader range of econometric problems where a group-structure exist, and the prior can accommodate different types of group structures.

varying components along with group-shrinkage helps exploiting the high-frequency data effectively as they get released. During the Covid-19 pandemic, the inclusion of t-distributed errors helps to channel the group-shrinkage towards relying more heavily on indicators that capture sentiment and activity in the service sector, as well as indicators on housing, both of which reflected disruptions from economic lockdowns. By contrast, models with other priors profit less from including the time-varying components. The horseshoe prior reads information in a dense and diffuse way across indicators and nowcast periods, and the spike-and-slab prior select sub-sets of indicators but shifts less effectively between them.

The proposed model combines various features and nests existing models in the literature. MIDAS regressions are popular since they are computationally less demanding than mixed-frequency state space representations (Bai et al., 2013), and are able to exploit heterogeneous signals from indicators since they do not rely on their co-movement. Non-linear and non-parametric MIDAS structures have been used to address non-linearities in the link between indicators and target (Guérin and Marcellino, 2013; Ghysels et al., 2020). Instead, we focus on modeling non-linear features in the target variable via a time-varying trend and stochastic volatilities that account for extreme observations, since these have been found to improve predictive performance in a range of unobserved component models, VARs, and dynamic factor models (Stock and Watson, 2009; Clark, 2011; D'Agostino et al., 2013; Berger et al., 2016; Carriero et al., 2016; Antolin-Diaz et al., 2017). Our model is closest to Carriero et al. (2015) who combine a MIDAS structure with stochastic volatility, but standard Minnesota type shrinkage. Antolin-Diaz et al. (2021) allow for a time-varying trend and stochastic volatility, exploit higher-frequency indicators via a dynamic factor model, and model outliers via an additive component in GDP growth. Instead, we adopt a Student's t-distribution to the shock structure, which can incorporate both smaller and more frequent variance shifts as well as rare extreme outliers (Jacquier et al., 2004; Clark and Ravazzolo, 2015).

The GIGG prior can be seen as an extension of the popular horseshoe prior (Carvalho et al., 2010) towards three-tiered group shrinkage. Other grouped shrinkage priors, such as Laplace-type (Casella et al., 2010) or Cauchy regularization (Xu et al., 2016), assume non-exchangeability in a group, but apply a uniform level of shrinkage within group, with no inference on the correlation via a covariate level scale. Mogliani and Simoni (2021) show that by augmenting Laplace-type regularization to be adaptive with spike-and-slab variable may provide improvements over the the former priors in grouped MIDAS regressions. Xu and Ghosh (2015), on the other hand, propose a group-sparse spike-and-slab prior which applies selection both on the group and within-group level, akin to the sparse group-lasso of Simon and Tibshirani (2012). Babii et al. (2022) study this estimator from a frequentist point of view in the context of MIDAS models and find that it provides substantial improvements for predictions. The GIGG prior can, due the flexibility of its three tiers, be thought of as a continuous generalization to these approaches that can flexibly adapt to any correlation structure and is therefore suitable for any basis of the MIDAS weighting function.

Finally, the ex-post sparsification approach relates to the debate of shrinkage versus sparsity initiated in Giannone et al. (2021) who highlight the importance of separately modeling shrinkage and sparsity. Our proposed methodology, too, separates shrinkage from the prior and sparsity from group selection and is able to provide uncertainty quantification in selection via inclusion probabilities over time. A key difference to Giannone et al. (2021) or group selection methods presented in Mogliani and Simoni (2021) and Babii et al. (2022), is that we view the proposed algorithm as a tool to understand and communicate the predictions of the model, coherent with decision theory, rather than to detect sparsity as a function of the actual observations of the target variable. At the same time, the GIGG prior can be tuned to incorporate the knowledge that macroeconomic data is highly correlated which facilitates inference when many small coefficients are likely, avoiding an “illusion of sparsity” (Giannone et al., 2021) that might otherwise be present with selection priors.

The remainder of the paper is structured as follows. Section 2 presents the T-SV-t BMIDAS, the GIGG prior and sparsification step. Section 3 outlines the data set and setup of the empirical application. Section 4 discusses nowcast evaluation results and the role of the model features, and section 5 discusses the results through the lens of variable inclusion probabilities. Section 6 concludes. The accompanying appendix shows details on methodology and additional results.

2 T-SV-t BMIDAS, group-shrinkage and sparsification

2.1 BMIDAS with time-varying components

The proposed T-SV-t BMIDAS combines time-varying component features with a multivariate MIDAS regression: (1) a time-varying trend component (2) stochastic volatility processes, and

(3) t-distributed errors in the the observation equation that allow for fat tails. Let y_t be the scalar valued target variable observed at discrete time points $t = (1, \cdots, T)$ which is modelled by unobserved states at frequency t and K covariates intermittently observed $m \geq 1$ times, $x_{t-(j-1)/m,k}$ for $j = (1, \cdots, m)$ and $k = (1, \cdots, K)$. In this paper, t will refer to quarters and $m = 3$ to months. Note that for ease of exposition, we choose m to be the same for all k, which will be relaxed following section 3. The model takes the following state-space form:

$$
\begin{aligned}
y_t &= \tau_t + \sum_{k=1}^{K} \mathcal{B}(L^{1/m}; \theta_k) x_{k,t} + \sqrt{\lambda_t} e^{\frac{1}{2}(h_0 + w_h \tilde{h}_t)} \tilde{\epsilon}_t^y, \\
\tilde{\epsilon}_t^y &\sim N(0,1),\ \lambda_t \sim IG(\nu/2, \nu/2)
\end{aligned} \tag{1}
$$

$$
\tau_t = \tau_{t-1} + e^{\frac{1}{2}(g_0 + w_g \tilde{g}_t)} \tilde{\epsilon}_t^\tau,\ \tilde{\epsilon}_t^\tau \sim N(0,1) \tag{2}
$$

$$
\begin{aligned}
\tilde{h}_t &= \tilde{h}_{t-1} + \tilde{\epsilon}_t^h,\ \tilde{\epsilon}_t^h \sim N(0,1) \\
\tilde{g}_t &= \tilde{g}_{t-1} + \tilde{\epsilon}_t^g,\ \tilde{\epsilon}_t^g \sim N(0,1),
\end{aligned} \tag{3}
$$

where (1) is the observation equation and (2)-(3) describe the evolution of states, latent trend and stochastic volatilities, respectively.

The trend component τ_t is modelled as a driftless random walk. This has long tradition in inflation trend estimation (see e.g. Stock and Watson (2007); Harvey et al. (2007)), but has recently also been introduced to nowcast models of GDP growth using dynamic factor models (Antolin-Diaz et al., 2017, 2021). The rationale for a unit root process in the trend, compared to the assumption of discrete changes in mean via structural breaks as in Kim and Nelson (1999) or McConnell and Perez-Quiros (2000) is to capture slow moving changes in GDP growth that may not necessarily be captured by the higher frequency components. This information typically relates to short run fluctuations of the business-cycle. [2] Nevertheless, which type of fluctuation the trend captures is not unequivocal, as the degree of smoothness of the latent trend depends on the priors employed and on the residual variation explained by the other model components (1) and (3). In the empirical application, we therefore assess the role of the trend by shutting down model components, and check the sensitivity of the trend to model specifications and priors.

Conditionally on the trend, a MIDAS component, $\sum_{k=1}^{K} \mathcal{B}(L^{1/m}; \theta_k) x_{k,t}$, relates short-term fluctuations from a possibly wide range of high-frequency indicators to y_t. Define $L^{1/m}$ as the back-shift

[2] A different non-stochastic approach to account for long-run growth features is presented in Giannone et al. (2019), where iterative forecasts are enforced to return to a long-run cointegrating equilibrium.

operator such that $L^{1/m}x_{t,k} = x_{t-1/m,k}$, and denote by θ a L-dimensional vector of coefficients. Then $\mathcal{B}(L^{1/m}, \theta_k)$ defines a lag polynomial, which, as is common in the MIDAS literature (e.g. Babii et al. (2022)), is parameterized by a weighting function, $\omega : \mathbb{R} \times \mathbb{R}^L \to \mathbb{R}$:

$$B(L^{1/m}, \theta_k)x_{t,k} = \sum_{j=1}^{qm} \omega(\frac{j-1}{qm}; \theta_k)x_{t-(j-1)/qm,k}. \tag{4}$$

Here, q indicates how many quarterly lags enter the polynomial in addition to the contemporaneous m months. Many different weighting functions have been proposed in the literature (Ghysels et al., 2007). In this paper, we consider ω to be a linear functional, $\omega(\theta_k) = \sum_{l=1}^{L} \theta_{k,l}\omega_{l,j}$ for $l = (1, \cdots, L)$ of which unrestricted weights (U-MIDAS, see Ghysels et al. (2007); Foroni et al. (2015)), and restricted Almon lag polynomial weights (Almon, 1965; Ferrara et al., 2022) are popular subsets thereof [3,4]. While U-MIDAS treats each higher frequency lag polynomial as an independent covariate, Almon polynomials force the curvature of regression coefficients belonging to a given covariate k to adhere to a L-degree polynomial process. This has particular appeal for macro time-series applications since economically relevant restrictions to the polynomial's end-points can be enforced, such as constraining the functional to peter out smoothly to zero for distant lags (Smith and Giles, 1976). Denote these restrictions by $r \forall k$. While the restrictions may induce parsimony when $L-r << qm$, the implied linear transformations typically induce a highly collinear grouping structure that can cause mixing problems for shrinkage priors (Griffin and Brown, 2012). The importance of modelling this correlation will become apparent when comparing priors that assume within lag group exchangeability to the GIGG prior and its sparsification.

Finally, we allow for two types of residual variation: (1) stochastic volatilities, $\{\tilde{h}_t, \tilde{g}_t\}$, for GDP growth, y_t, and the latent trend, τ_t, that model persistence in error variance, and (2) student-t errors in the y_t equation, to model non-persistent shocks such as extreme events and outliers that are otherwise unaccounted for by the other model components (Jacquier et al., 2004; Clark and Ravazzolo, 2015). Since states $\{h_t, g_t\}$ can be regarded as parameters, the state components alone can quickly make the model high dimensional. To allow for stronger shrinkage on the states

[3] Foroni and Marcellino (2014) show that linear MIDAS methods are competitive with non-linear MIDAS weighting schemes such as the non-linear Almon and beta functions (Ghysels et al., 2004, 2007; Andreou et al., 2010; Ghysels et al., 2020), while being compatible with off the shelf shrinkage methods. Linear MIDAS have been employed in (Foroni et al., 2015; Angelini et al., 2011; Mogliani and Simoni, 2021).

[4] In general, however, any arbitrary function can be used for ω which Babii et al. (2022) denote as the dictionary of functions.

via priors on w_h and w_g, which control state-smoothness, we use the non-centered state space formulation after Frühwirth-Schnatter and Wagner (2010). This retains conditional conjugacy with the priors discussed below.[5] The introduction of t-errors may benefit density nowcasts since fat-tails imply wider probability bands around extreme observations, but can also be relevant for point nowcasts because t-distributed errors discount large contemporaneous movements in y_t and, therefore, limit the propagation of outliers to the posteriors of the model components (Chiu et al., 2017). Differently from linear outlier treatments on the level of the target variable (Stock and Watson, 2016; Antolin-Diaz et al., 2021), t-distributed errors reflect a continuous range of time variation and can incorporate both smaller, more frequent variance shifts and rare extreme outliers.

2.2 Priors

We now describe the prior hierarchy used for drawing inference on all unknowns in (1)-(3):

$$\pi(\zeta) = \pi(\theta)\pi(\boldsymbol{\tau})\pi(\tilde{\boldsymbol{h}})\pi(\tilde{\boldsymbol{g}})\pi(\phi)\pi(\boldsymbol{\lambda}|\nu)\pi(\nu), \tag{5}$$

where ζ collects all unknowns into one vector. The MIDAS coefficients θ are regularized by the GIGG prior with group-shrinkage structure. Bold-faced letters refer to time-ordered vectors (e.g., $\boldsymbol{\tau} = (\tau_1, \cdots, \tau_T)'$), and ϕ collects any remaining state parameters $(\tau_0, h_0, g_0, w_h, w_g)'$. $\{\tau_0, h_0, g_0\}$ are the starting values for the latent states.

2.2.1 GIGG prior on MIDAS coefficients

Multivariate MIDAS regressions are highly parameterized, but the grouping and correlation structure across coefficients can be exploited for efficient shrinkage. We do so in three ways: shrinkage across all coefficients, group-wise shrinkage jointly for coefficients belonging an indicator, and modelling of the correlation between coefficients within a group. The prior, therefore, speaks to the fact that, first, not all of the k indicators are equally relevant for prediction, addressed the *group-wise shrinkage* of lags belonging to a given indicator. Second, correlation among parameters *within the lag group* is typically very high for consecutive high-frequency lags, particularly when functional

[5] Since w_g and w_h appear in the observation and state equation, one can apply conjugate normal priors which exert stronger shrinkage than the inverse-gamma priors conventionally applied in Bayesian state-space models. This approach has been used in large time-varying parameter models, see Huber et al. (2021). The more commonly employed centered SV process h_t can be exactly recovered given that $h_t = h_0 + w_h \tilde{h}_t$ (likewise for g_t).

restrictions such as those implied by the Almon polynomial are applied. Exchangeable priors that leave this unaddressed, are liable to random covariate selection and bad mixing (Boss et al., 2021; Piironen et al., 2020). Third, group-structure and correlation across lags can *jointly* matter for predictive performance since the relative impact of the lag group on the target variable can be affected. The prior, "inverse-Gamma Gamma" (GIGG) (Boss et al., 2021), is specified as

$$\begin{aligned} \theta_{k,j} \sim N(0, \vartheta^2 \gamma_k^2 \varphi_{k,j}^2), \ \ \forall j \in \{1, \cdots, L-r\} \\ \vartheta \sim C_+(0,1), \ \ \gamma_k^2 | a_k \sim G(a_k, 1) \ \ , \varphi_{k,j}^2 \sim IG(b_k, 1), \end{aligned} \tag{6}$$

where $G(\bullet, \bullet)$, $IG(\bullet, \bullet)$ and $C_+(\bullet, \bullet)$ refer to the Gamma, inverse-Gamma, and half Cauchy distribution with positive support, respectively. ϑ controls the overall level of sparsity. γ_k acts as a shrinkage factor on the joint impact of coefficients from group k, and $\varphi_{k,j}$ controls how correlated group members are within k. The magnitude and relative size of a_k and b_k determine the amount of shrinkage and relative importance between group-shrinkage and within-group correlation.

Depending on the choice of hyperparameters, prior (6) is related to several previous approaches employed in MIDAS settings. For a group-size of 1 and setting $a_k = b_k = 0.5$, it reduces to the horseshoe prior of Carvalho et al. (2010) which Kohns and Bhattacharjee (2022) apply to nowcasting US GDP growth with U-MIDAS sampled components. The GIGG can also be viewed as a continuous generalization to the group adaptive lasso with spike-and-slab group selection which Mogliani and Simoni (2021) analyze empirically and theoretically for Almon MIDAS regressions. Instead of Laplacian slab distributions, however, the GIGG allows for fatter tailed distributions and inference on the correlation of coefficients within a group. More broadly, for any group-size larger than 1, the prior follows a correlated normal beta prime distribution akin to a grouped formulation of the prior by Armagan et al. (2013), $\gamma_k^2 \varphi_{k,j}^2 \sim \beta'(a_k, b_k)$ which flexibly nests the group-horseshoe Xu et al. (2016)[6] and may also be viewed as a continuous approximation to the group-sparse spike-and-slab prior of Xu and Ghosh (2015).

For the application to nowcasting, we leverage the model structure to inform on the hyperparameters of the GIGG prior. Generally, the lower a_k, the stronger the group is penalized, while the higher b_k, the more within-group correlation is allowed for. But the relative magnitudes matter as well. To visualize these relationships, Figure 1 shows the prior distribution on the shrinkage

[6] Unlike Xu et al. (2016), the prior reduces to the exact horseshoe at group-size 1.

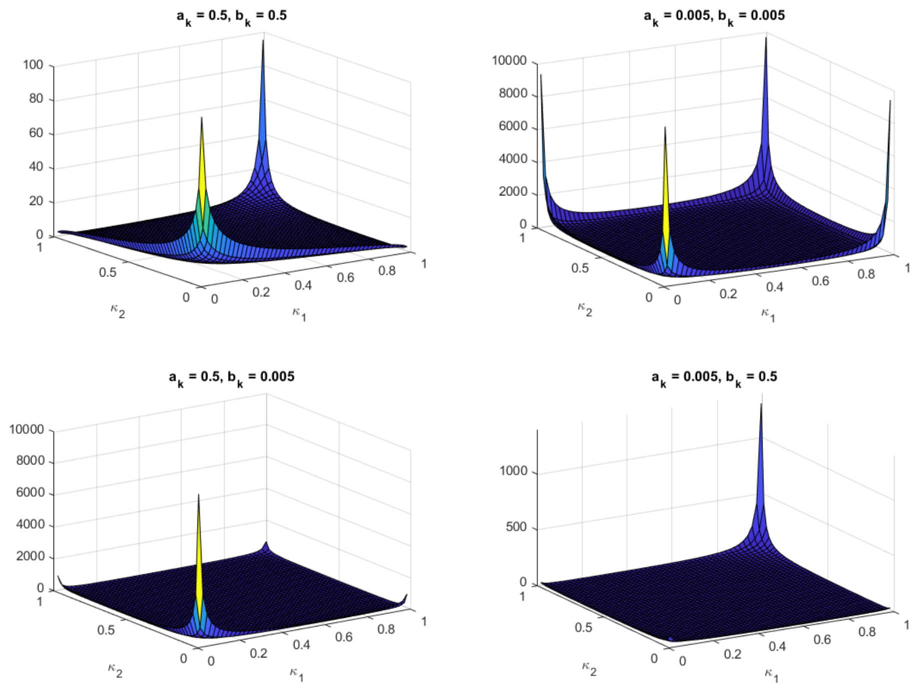


Figure 1: Bi-variate shrinkage coefficient plots for various hyper-parameter values. a_k controls group-level shrinkage while b_k controls the degree of correlation within groups.

coefficients $\kappa_j \in [0, 1]$, implicitly defined by $(\vartheta, \gamma_k, \varphi_{k,j})$, for different choices for a_k and b_k and a group of 2. These are often analyzed in the context of generalized linear models (Piironen et al., 2017) since they inform on how much the prior shrinks toward a mean of 0 from the maximum likelihood estimate of the regression weights (the accompanying appendix shows full derivations the MIDAS component). High probability mass on 1 implies aggressive shrinkage toward 0 and the degree of concentration of the joint probability distribution informs on the degree of correlation in shrinkage.

Two general shrinkage profiles emerge from choosing $a_k = b_k$ or $a_k \neq b_k$. Setting $a_k = b_k$ induces a horseshoe type shape, which displays high correlation when set to $a_k = b_k = 0.5$ (upper left plot). The high degree of correlation favors to either jointly shrink the group members to 0, or leave the coefficients relatively un-shrunk, akin to the group-horseshoe by Xu et al. (2016). For smaller values such as 0.005 (upper right plot), on the other hand, we observe the tendency for independent horseshoe type shrinkage, implying low correlation. A low magnitude of a_k and b_k can be appropriate when prior knowledge exists that only selected lags are important. When $a \neq b$, within-group correlation is high and the horseshoe shape collapses towards very high or very little joint shrinkage for *both* coefficients. Choosing $a >> b$ (lower left sub-plot) favors little regularization (high mass at South edge). Choosing $a << b$ (lower left sub-plot) induces aggressive joint shrinkage (high mass at North edge), and the overall impact of both coefficients is pushed toward sparsity. Nonetheless, in both cases, small mass in the edges retains the flexibility to escape the joint shrinkage tendencies.

For our empirical application, we adopt this latter case of aggressive group-correlated shrinkage, setting $a = 1/T, b = 0.5$ for all coefficient groups. This choice reflects the view that a high degree of correlation is present between MIDAS lag coefficients belonging to the same indicator, particularly when coefficients are restricted via Almon polynomials. And a comparison across different hyperparameter choices suggests an improved nowcasting performance with strong prior correlation in shrinkage, particularly prior to the pandemic.

2.2.2 Priors for the latent states and other coefficients

For priors related to the states and starting conditions, we follow the previous literature. For the latent states $(\boldsymbol{\tau}, \tilde{\boldsymbol{h}}, \tilde{\boldsymbol{g}})$ we consider a joint normal prior derived using methods proposed in Chan and Jeliazkov (2009). This allows representing the entire conditional state posterior as a tractable normal distribution. Latent states are prone to over-fitting in over-parameterized models and the smoothness of the state processes are notoriously difficult to identify from the data alone (Cogley et al., 2005; Antolin-Diaz et al., 2017). We therefore put normal priors on the state standard deviations with small prior variance that exert stronger shrinkage than commonly employed inverse-Gamma priors (Frühwirth-Schnatter and Wagner, 2010), $w_i \sim N(0, 0.01^2)$ for $i \in \{h, g\}$. This encourages the states to be slow moving. We set weakly informative priors on the initial conditions $\{i_0\}_{i\in\{h,g,\tau\}} \sim N(0, 10)$. Lastly, we put a relatively uninformative uniform prior on the degrees of freedom of the t-distributed errors in the observation equation, $\nu \sim \mathcal{U}[2, 50]$. The lower bound is chosen to insure the existence of the first two moments of the t-distribution (Chan and Grant, 2016).

2.3 Estimation Algorithm

Here, we briefly review the estimation algorithm, firstly defining the likelihood stacked likelihood across all observation. Let 1_T be a column of ones. Recall that $\boldsymbol{h} = h_0 1_T + w_h \tilde{\boldsymbol{h}}$ and $\boldsymbol{g} = g_0 1_T + w_g \tilde{\boldsymbol{g}}$

and define $\boldsymbol{\tau} = (\tau_1, \cdots, \tau_T)$. Then it follows that

$$(\boldsymbol{y}|\boldsymbol{X}, \theta, \boldsymbol{\tau}, \tilde{\boldsymbol{h}}, h_0, w_h, \boldsymbol{\lambda}) \sim N(\boldsymbol{Z}\theta + \boldsymbol{\tau}, \Lambda_h \Lambda_\lambda), \tag{7}$$

where $\boldsymbol{Z}\theta$ are the vectorized MIDAS coefficients and $\Lambda_h = diag(e^{h_1}, \cdots, e^{h_T})$ and $\Lambda_\lambda = diag(\lambda_1, \cdots, \lambda_T)$.[7] The assumption of independence between MIDAS coefficients and state space components allows to derive conditional posterior distributions via an efficient Metropolis-within-Gibbs sampling algorithm. We sequentially simulate from the following algorithm:

1. Sample $\theta \sim p(\theta|\boldsymbol{y}, \bullet)$
2. Sample hyper-parameters $(\vartheta, \gamma_k^2, \varphi_{k,j}^2, \nu_\vartheta)$ in one block
   (a) $\vartheta^2 \sim p(\vartheta^2|\boldsymbol{y}, \bullet)$
   (b) $\gamma_k^2 \sim 1/p(\gamma_k^{-2}|\boldsymbol{y}, \bullet)$ for for $k = (1, \cdots, K)$
   (c) $\varphi_{k,j}^2 \sim p(\varphi_{kj}^2|\boldsymbol{y}, \bullet)$ for $j = (1, \cdots, L - r)$
   (d) $\nu_\vartheta \sim p(\nu_\vartheta|\boldsymbol{y}, \bullet)$
3. Sample $\tilde{\boldsymbol{\tau}} \sim p(\tilde{\boldsymbol{\tau}}|\boldsymbol{y}, \bullet)$ and $\tau_0 \sim p(\tau_0|\boldsymbol{y}, \bullet)$
4. Sample $\tilde{\boldsymbol{h}} \sim p(\tilde{\boldsymbol{h}}|\boldsymbol{y}, \bullet)$, $h_0 \sim p(h_0|\boldsymbol{y}, \bullet)$ and $w_h \sim p(w_h|\boldsymbol{y}, \bullet)$
5. Sample $\tilde{\boldsymbol{g}} \sim p(\tilde{\boldsymbol{g}}|\boldsymbol{y}, \bullet)$, $g_0 \sim p(h_0|\boldsymbol{y}, \bullet)$ and $w_g \sim p(w_g|\boldsymbol{y}, \bullet)$
6. Sample $\lambda_t \sim p(\lambda_t|\boldsymbol{y}, \bullet)$ for $t = (1, \cdots, T)$
7. Sample $\nu \sim p(\nu|\boldsymbol{y}, \bullet)$ with a Metropolis step.

ν_ϑ is an $IG(\bullet, \bullet)$ distributed mixture variable that, when $\vartheta^2|\nu_\vartheta \sim IG(1/2, 1/\nu_\vartheta)$ and $\nu_\vartheta \sim IG(1/2, 1)$, gives exact draws for $(\vartheta|\boldsymbol{y}, \bullet)$ (Makalic and Schmidt, 2015). We iterate sampling steps 1.-7. 5000 times for burn-in and retain further 5000 samples for inference.[8] To speed up the computations of the posteriors, we make use of the state sampling techniques of Chan and Jeliazkov (2009) for the state variables. This allows drawing steps 3.-5. in a non-recursive fashion, which increases sampling and computational efficiency via sparse-matrix operations compared to Kalman filter based techniques (Carter and Kohn, 1994). We sample from the posteriors of $(\tilde{\boldsymbol{h}}, \tilde{\boldsymbol{g}})$ using the

[7] Define $X_k = (x_{k,t-\frac{j-1}{qm}})_{t\in\{1,\cdots,T\},j\in\{1,\cdots,qm\}}$ as a $T \times qm$ matrix of high frequency lags $\forall k$ and $W = (w_1', \cdots, w_{L-r}')$ as a $qm \times L - r$ matrix of Almon restricted weights. Then, $Z_k = X_k W$ and the stacked matrix is $\boldsymbol{Z} = (Z_1, \cdots, Z_T)$ and the MIDAS functional coefficient vector is $\theta = (\theta_1', \cdots, \theta_K')'$.

[8] To check convergence, we tested with 20000 iterations for burn-in and 20000 for inference, and results remain similar.

approximate sampler of Kim et al. (1998). The conditional MIDAS coefficient posterior is normal and we use the algorithm of Bhattacharya et al. (2016) to speed up computation and aid mixing when $\sum_{k=1}^{K}(L-r) >> T$. The accompanying appendix sets out details on the conditional posterior distributions and the Metropolis step.

2.4 Group-sparsification on the posterior

Continuous priors such as the GIGG may already achieve optimal empirical and prediction risk properties (Chakraborty et al., 2020; Boss et al., 2021). However, when the policy maker aims to understand and communicate nowcasts, the fact that with such priors, posteriors remain non-zero can be a limitation since the impact on nowcasts remains opaque. In view of this, we propose to *ex-post* sparsify the posterior of the regression weights based on decision theory, with the aim of finding the smallest subset of indicators which achieve closest predictive performance to the unsparsified model. Compared to priors that conduct selection, we view sparsity as a decision tool that is separate from any regularization that the prior imposes. Giannone et al. (2021) highlight the importance of separating shrinkage from sparsity, particularly for macroeconomic data, where high correlation among indicators can result in falsely detecting sparsity ("illusion of sparsity").

We derive a new analytical solution for ex-post sparsification for the general MIDAS posterior which takes the temporal grouping structure of the MIDAS functional into account. Specifically, we threshold the coefficients of lags pertaining of an indicator jointly to zero if they have little effect on the predictions. And we use insights from the sparse group lasso literature (Simon and Tibshirani, 2012; Breheny and Huang, 2015) to formulate a computationally efficient solution to the group selection problem which is robust to the intra-group correlation, and applicable to any MIDAS specification using linear weights for (4).

Suppose we take the perspective of a decision maker who minimizes a utility function over the Euclidean distance between a linear model that penalizes group-size akin to Zou (2006) and the predictions from model (1)-(3).

$$\mathcal{L}(\tilde{\boldsymbol{Y}}, \alpha) = \frac{1}{2}||\boldsymbol{Z}\alpha - \tilde{\boldsymbol{Y}}||_2^2 + \sum_{k=1}^{K} \phi_k ||\alpha_k||_2, \tag{8}$$

where $\tilde{\boldsymbol{Y}}$ refers to a realization from the posterior predictive distribution $p(\tilde{\boldsymbol{Y}}|\boldsymbol{y}) =$

$\int p(\tilde{\boldsymbol{Y}}|\boldsymbol{y}, \theta, \bullet)p(\theta|\boldsymbol{y}, \bullet)d\theta$, and $||\bullet||_p$ refers to the ℓ_p-norm.[9] α is an unknown MIDAS weight vector which due to the group-wise penalty $\sum_{k=1}^{K} \phi_k ||\alpha_k||_2$ may display sparsity via penalty parameter ϕ_k for each indicator group. Similar to the logic of adaptive group-lasso (Wang and Leng, 2008), the penalization term induces non-differentiability at zero, which creates a soft-thresholding effect between $[-\phi_k, \phi_k]$, thereby forcing the coefficients on all group members to zero. Unlike traditional selection approaches such as the spike-and-slab prior, sparsity is induced using the prediction of $\boldsymbol{Y}$ as a target, rather than $\boldsymbol{Y}$ directly which has been previously shown to benefit the stability of subset selection (Piironen and Vehtari, 2017).

The Bayes optimal solution for α is obtained by integrating out the two sources of uncertainty which characterize the uncertainty in $\tilde{\boldsymbol{Y}}$: posterior uncertainty from the predictive distribution, as well as in the parameters θ (Lindley, 1968). Here, we show the analytical solution for (8) and discuss the assumptions needed to derive it (for a full derivation, see appendix). The sparsified estimate $\alpha_k^{*(s)}$ for each Gibbs-sampling step $s = 1, \cdots, S$, is given by:

$$\alpha_k^{*(s)} = \left(||\theta_k^{(s)}||_2 - \phi_k^{(s)} \right)_+ \frac{\theta_k^{(s)}}{||\theta_k^{(s)}||_2}, \ \forall k \in \{1, \cdots, K\}, \tag{9}$$

where $(x)_+ = \max(x, 0)$. (9) implies that when $\theta_k^{(s)}$ are close to $\mathbf{0}$, then $\alpha_k^{*(s)} = \mathbf{0}$, whereas, when $\theta_k^{(s)}$ are sufficiently large, then $\alpha_k^{*(s)} = (1 - \frac{\phi_k^{(s)}}{||\theta_k^{(s)}||_2^2})\theta_k^{(s)}$, in which case the first term will be very close to 1, thus imposing close to no further shrinkage. Two assumptions are needed to derive (9). Firstly, it requires orthonormalisation of the data for each k such that $T^{-1}\tilde{\boldsymbol{Z}}_{\boldsymbol{k}}'\tilde{\boldsymbol{Z}}_{\boldsymbol{k}} = I_{L-r}$. This serves to simplify the make the solution robust to any correlation implied by ω within (4). Further, as shown in Simon and Tibshirani (2012), not orthonormalizing groups ignores the cross-correlation of group members in k, in which case the algorithm implicitly prefers to not threshold groups with large covariance, and ignores that $\boldsymbol{Z}_k$ might have different scales.[10] Secondly, we make use of work by Ray and Bhattacharya (2018) and Chakraborty et al. (2020) who show that, when setting $\phi_k^{(s)} = \frac{1}{||\theta_k^{(s)}||_2}$, iterative solution methods such as the coordinate descent (Friedman et al., 2010) converge after the first cycle.

We then quantify (9) the posterior "inclusion probability" that a given group of coefficients was

9 Note that for simplicity we define the predictive distribution over in-sample values of $\boldsymbol{Z}$, but in principle any data can be used for the analysis.

10 It can be further shown that orthonormalizing the objective, establishes connection to best subset selection and uniformly most powerful invariant testing (Simon and Tibshirani, 2012)

included in the model's forecasts:

$$p(\alpha_k^{*(s)} \neq 0|\tilde{\boldsymbol{Y}}) \approx \frac{1}{S}\sum_{s=1}^{S} \mathbb{1}_{\alpha_k^{*(s)} \neq 0} \; \forall k, \tag{10}$$

and associated posterior variance: $var(\alpha_k^{*(s)} \neq 0|\tilde{\boldsymbol{Y}}) \approx \frac{1}{S}\sum_{s=1}^{S}(\mathbb{1}_{\alpha_k^{*(s)} \neq 0} - \overline{\mathbb{1}}_{\alpha_k \neq 0})^2$. Thus, posterior inclusion probabilities represent the relative frequency with which the group k was selected in the sparsified estimate $\alpha^{*(s)}$ over all Gibbs draws. These reflect the uncertainty around the inclusion of a variable and its relative impact, which would not be the case if we were to sparsify only once on the posterior mean.[11]

Linking regularization from priors and variable selection has long tradition in statistics (Lindley, 1968). Hahn and Carvalho (2015) re-popularized this approach for high-dimensional problems by introducing independent lasso-type penalties into the decision maker's utility function. Ray and Bhattacharya (2018) and Chakraborty et al. (2020) propose computationally efficient solutions for this decision task for individual covariate and group-selection, respectively. These have been adapted to time-varying parameter models and VARs (Huber et al., 2021; Hauzenberger et al., 2021), where Huber et al. (2021) also sparsify over MCMC draws and derive posterior inclusion probabilities. But to our knowledge, existing applications to MIDAS settings have assumed independent covariate selection (Kohns and Bhattacharjee, 2022). And differently from the previous approaches, the group-sparse solution is robust to correlation within group and therefore generalizes to any linear MIDAS approach.

We note at this point, that our sparsification approach is no panacea for detecting true sparsity of the data generating process. Which indicators are ultimately thresholded to zero, depends on the prior applied for the model and its ability to apply aggressive shrinkage to variables that are weakly related to the target, y_t. To highlight the importance for group shrinkage in MIDAS regression, we will apply our sparsification algorithm also to previously used priors for macro settings which assume exchangeability within group such as the horseshoe prior and spike-and-slab. While Giannone et al. (2021) conclude that the data are typically non-informative about the which subset of predictors to include in the model, we show that this is not necessary the case

[11] See Woody et al. (2021) and Chakraborty et al. (2020) for formal justification of the model selection uncertainty and the asymptotic risk properties, respectively. Note, that unlike spike-and-slab priors, we do not conduct direct inference on the inclusion probability of a variable, hence deriving its approximate posterior intervals are not immediately available.

when imposing three-tiered group shrinkage and then sparsifying ex-post, which in our application helps us communicating intuitive sub-sets of most relevant predictors over time.

3 Empirical setup

For our application, we nowcast real quarter-on-quarter GDP growth in the United Kingdom based on a set of monthly macroeconomic indicators following a stylized publication calendar. In the following, we outline the data set, the stylized publication calendar we follow, and the set-up of our nowcasting exercise and evaluation.

3.1 Data set

The set of monthly macroeconomic indicators has been compiled to reflect information on the UK economy that policymakers actively monitor to gauge economic activity in real time, and is comparable to data sets employed in previous studies (Antolin-Diaz et al., 2017; Anesti et al., 2018). We include a range of real activity and survey indicators, including indices of production and services, exports and imports, a range of labor market series, as well as timely business and consumer surveys (CBI survey, PMIs, GFK). In order to capture lending conditions that can affect economic conditions via financial conditions we also include mortgage lending approvals and VISA credit card consumer spending. These series also tracked consumer spending during the pandemic, reflecting shut-downs of business and housing activity.[12] We do not add asset prices or other financial indicators which have been found to contribute little to nowcast updates once information from monthly survey and real activity data is accounted for (Bańbura et al., 2013; Anesti et al., 2018). Also, during the Covid-19 period financial markets were detached from real activity in the UK—asset prices initially collapsed, then stabilized early on in the pandemic in response to monetary policy interventions, and subsequently exhibited a boom that was not in line with the weakness of the real economy. The series are transformed to be approximately stationary

[12] Alternatively, various studies have exploited new data sources at a daily or weekly frequency (Ng, 2021; Baumeister et al., 2021; Huber et al., 2020; Kapetanios et al., 2022). Including such data faces the challenge of very short time-series, and we therefore abstain from including them in our main analysis. We have, however, explored the possibility of linking the monthly VISA consumer spending data with experimental weekly debit and credit card data provided by the UK's Office for National Statistics. This analysis shows additional nowcast gains in the first weeks of the reference quarter, which however dissipate once the monthly series becomes available. Result for this exercise are available upon request.

prior to estimation.[13]

We consider the sample period from 1999Q1 to 2021Q3. The start of the sample is pinned down by data availability since many of the monthly indicators are not available for earlier years.[14] To mimic incoming information over the data release cycle that a nowcaster would face in reality, we produce nowcasts based on a pseudo real time data calendar, as outlined below. However, for ease of analysis we use final vintages of the data, downloaded in December 2021. Since our focus here lies in understanding the proposed model, we leave an account for real time data releases and revisions for future research.[15]

3.2 Nowcast exercise

Macroeconomic data are published asynchronously at different points in time and with delays ranging from various weeks (survey data) to up to various months (labour market data) after the reference month. To simulate the information set available to the nowcaster over the data release cycle, we follow a stylized pseudo real-time data release calendar (see Table 1).

As is common with MIDAS approaches, we start the nowcast exercise for each quarter anew. We start predicting with all available information on the first of the month of the reference quarter. Following the stylized release calendar in Table 1, we generate overall 20 nowcasts that are being produced at each date in the quarter when new data series are typically released, until the release of quarterly GDP six weeks after the end of the reference quarter.[16] For each new data release over the data release cycle, we generate nowcasts from the predictive distribution $p(y_{t+1}|\Omega_T^{\nu})$, where $v = 1, \cdots, 20$ refers to the nowcast periods and Ω_T^v represents the information set that expands with each data release. Since the MIDAS framework belongs to the class of reduced-form mixed frequency models, each information set Ω_T^v results in a different model, depending on which data are observable over the data release cycle (Carriero et al., 2015).To draw samples from the predictive

13 The appendix includes an overview of the data and their respective transformations.

14 Some of the series have missing values at the beginning of the sample period. We interpolate these based on a principal component (PCA) model that accounts for missing information via the alternating least square algorithm, and results also remained similar when employing the Expectation Maximization algorithm (Bańbura and Modugno, 2014) instead.

15 See Anesti et al. (2018) for and analysis of UK data on the forecastability of different vintages and how to incorporate that information for nowcast updates.

16 We refer to the first GDP publication, available about 40 days after the reference quarter. We abstain from accounting for a less accurate preliminary GDP estimate that was available 25 days after the reference quarter prior to Sir Charles Bean's 2018 review of UK economic statistics (Scruton et al., 2018).

Table 1: Stylised pseudo real-time data release calendar.

Nowcast	Quarter	Days to GDP	Month	Timing within month	Release	Publication Lag
1		135	1	1st of month	PMIs	m-1
2		125	1	End of 2nd week	IoP, IoS, Ex, Im	m-2
3		120	1	3rd week	Labour market data	m-2
4		115	1	3rd Friday of month	Mortgage & Visa	m-1
5		110	1	End of 3rd week	CBIs & GfK	m
6	Reference	105	2	1st of month	PMIs	m-1
7	quarter	97	2	Mid of 2nd week	Quarterly GDP	q-1
8	(nowcast)	95	2	End of 2nd week	IoP, IoS, Ex, Im	m-2
9		90	2	3rd week	Labour market data	m-2
10		85	2	3rd Friday of month	Mortgage & Visa	m-1
11		80	2	End of 3rd week	CBIs & GfK	m
12		75	3	1st of month	PMIs	m-1
13		65	3	End of 2nd week	IoP, IoS, Ex, Im	m-2
14		60	3	3rd week	Labour market data	m-2
15		55	3	3rd Friday of month	Mortgage & Visa	m-1
16		50	3	End of 3rd week	CBIs & GfK	m
17		45	1	1st of month	PMIs	m-1
18	Subsequent	35	1	End of 2nd week	IoP, IoS, Ex, Im	m-2
19	quarter	30	1	3rd week	Labour market data	m-2
20	(backcast)	25	1	3rd Friday of month	Mortgage & Visa	m-1

Notes: "Timing" refers to typical data release times as of December 2021, abstaining from changes in the publication calendar over the sample period. "Release" refers to the data series updated at a given nowcast. "Publication lag" represents the delay relative to the reference quarter (i.e. publication at any point in the subsequent month considered to be one month lag, m-1). The appendix contains a list of data series included.

distribution, we integrate over all parameter uncertainties which is easily implemented via Monte Carlo integration (Cogley et al., 2005).

We start the nowcast exercise with an in-sample period of 1999Q1-2011Q1, and iteratively expand it until the end of the forecast sample, T_{end} = 2021Q3. Since the Covid-19 pandemic represents a historic shock to the macroeconomy, we separately evaluate nowcasts over a sample that ends in 2019Q4 and one that cover the full sample period including the Covid-19 shock.

Point nowcasts are computed as the mean of the posterior predictive distribution and are compared via root-mean-squared-forecast-error (RMSFE) calculated at each nowcast period as:

$$\text{RMSFE} = \sqrt{\frac{1}{T_{end}} \sum_{t=1}^{T_{end}} (y_{T+t} - \hat{y}^v_{T+t|\Omega^v_{T+t-1}})^2}, \tag{11}$$

where $\hat{y}^v_{T+t|\Omega^v_{T+t-1}}$ is the mean of the posterior prediction for nowcast period v using information until $T+t-1$ and T is the initial in-sample length. Forecast density fit is measured by the mean continuous rank probability score (CRPS):

$$\text{CRPS} = \frac{1}{T_{end}} \sum_{t=1}^{T_{end}} \frac{1}{2} \left| y_{T+t} - y^{v}_{T+t|\Omega^{v}_{T+t-1}} \right| - \frac{1}{2} \left| y^{v,A}_{T+t|\Omega^{v}_{T+t-1}} - y^{v,B}_{T+t|\Omega^{v}_{T+t-1}} \right|. \tag{12}$$

with $y^{v,A,B}_{T+j|\Omega^{v}_{T+j-1}}$ independently drawn from the posterior predictive density $p(y^{v}_{T+1|\Omega^{v}_{T+j-1}}|y_T)$. The CRPS belongs to the class of strictly proper scoring rules (Gneiting and Raftery, 2007), and can be thought of as the probabilistic generalization of the mean-absolute-forecast-error. The objective in terms of predictive precision is to minimize both evaluation metrics.

4 Model features and empirical nowcast performance

In the following, we discuss the results, with a focus on how the time-varying components and the group-shrinkage prior, respectively and jointly, contribute to the nowcast performance. In section 4.1, we compare the proposed T-SV-t-BMIDAS model with alternative models where we shut down time-variation, and in section 4.2, and we consider alternative priors on the MIDAS components for the model with and without time-varying components.

4.1 Role of time-varying unobserved components

Figure 2 looks at the role of the time-varying trend and stochastic volatilities with t-distributed errors by comparing the T-SV-t BMIDAS model with trend and stochastic volatility with t-distributed errors with the following alternatives, as well as an AR(2) benchmark (yellow line).

- *T-SV*: time-varying trend and stochastic volatility, normally distr. errors.
- *T, Const Var*: time-varying trend with constant variance.
- *SV-t*: no trend, stochastic volatility with t-distributed errors.
- *SV*: no trend, stochastic volatility, normally distributed errors.
- *Const Var*: no trend, constant variance.

Prior to the pandemic (left panels), there are substantial improvements in point and density nowcast for most models against the AR(2), and in particular for the proposed T-SV-t model compared to alternative models. Adding either a time-varying trend or stochastic volatility to

Table 2: Nowcast Evaluation Results

Nowcast Periods	Evaluation pre-pandemic Average	6	13	18	Evaluation incl. pandemic period Average	6	13	18
	RMSFE				RMSFE			
AR(2) benchmark (abs. RMSFE)	0.42	0.42	0.42	0.42	11.45	11.4	11.47	11.48
T-SV-t BMIDAS, GIGG w/ Spars. (rel. RMSFE)	0.66***	0.68**	0.60*	0.51**	0.21***	0.27	0.11	0.10
Alternatives to T-SV-t (all BMIDAS, GIGG w/ Spars)								
T-SV	0.75***	0.81*	0.72	0.44**	0.21***	0.27	0.17	0.08
T, Constant variance	0.76***	0.84	0.68	0.47**	0.21***	0.28	0.16	0.07
No T, SV-t	0.78***	0.90	0.70	0.46**	0.21***	0.31	0.10	0.08
No T, SV	0.81***	0.91	0.77	0.46**	0.22***	0.31	0.15	0.08
No T, Const. var.	1.03***	1.01	1.00*	0.67	0.22***	0.27	0.19	0.09
Alternatives priors on MIDAS coefficients (all with T-SV-t)								
GIGG w/out Spars.	0.69***	0.71*	0.62	0.49*	0.21***	0.32	0.10	0.09
Horseshoe (HS)	0.81***	0.87	0.73	0.74	0.28***	0.28	0.23	0.24
Spike and Slab (SS)	0.76***	0.74**	0.72*	0.78	0.32***	0.38	0.25	0.25
Alternatives to BMIDAS (all with T-SV-t, GIGG w/ Spars)								
U-BMIDAS	0.71***	0.74**	0.65**	0.60**	0.24***	0.29	0.18	0.18
MF-DFM	0.67***	0.75	0.67**	0.63**	0.32***	0.33	0.32	0.34
Combination univar. MIDAS	0.68***	0.68**	0.67**	0.67**	0.36***	0.37	0.34	0.33

Notes: The table shows the average RMSFE values for the AR model in the first row of each panel across all 20 nowcast periods ("Average"), and for selected nowcast periods (6,13,18). RMSFE values for the other models are in relative terms to the AR model and stars indicate significance as per the Diebold-Mariano test Diebold et al. (1998) ($*$ = 10% significance,**=5% significance,***=1% significance). T stands for time-varying trend, SV-t stands for stochastic volatility with t-distributed errors.

the model improves point and density forecasts significantly by about 20% relative to the AR(2) benchmark, on average across nowcast periods, as shown in Table (2) (RMSE results) and Table B2 in the appendix (CRPS results). Over the data release cycle, improvements of these models are only significant late in the data release cycle, however. Combining both the time-varying trend and SV-t in the proposed model leads to a stronger improvement by 35% on average, and also stabilizes performance over the data release cycle, with significant improvements throughout. By contrast, the model without trend and with constant volatility shows a volatile performance across nowcast periods and does not improve against the benchmark. This underlines that incorporating at least one, and preferably both of the proposed time-varying components is important to exploit high-frequency information with BMIDAS models, in line with existing evidence based on other models for the United States (Antolin-Diaz et al., 2017; Carriero et al., 2015).

When including the Covid-19 pandemic (right panels), nowcast errors are higher for all models, particularly early on in the data release cycle, and differences across model variants are relatively smaller. Nowcast performance of all models clearly improves with the release of "hard" indicators for the first months of the reference quarter (nowcast period 13, i.e. 65 days prior to GDP),

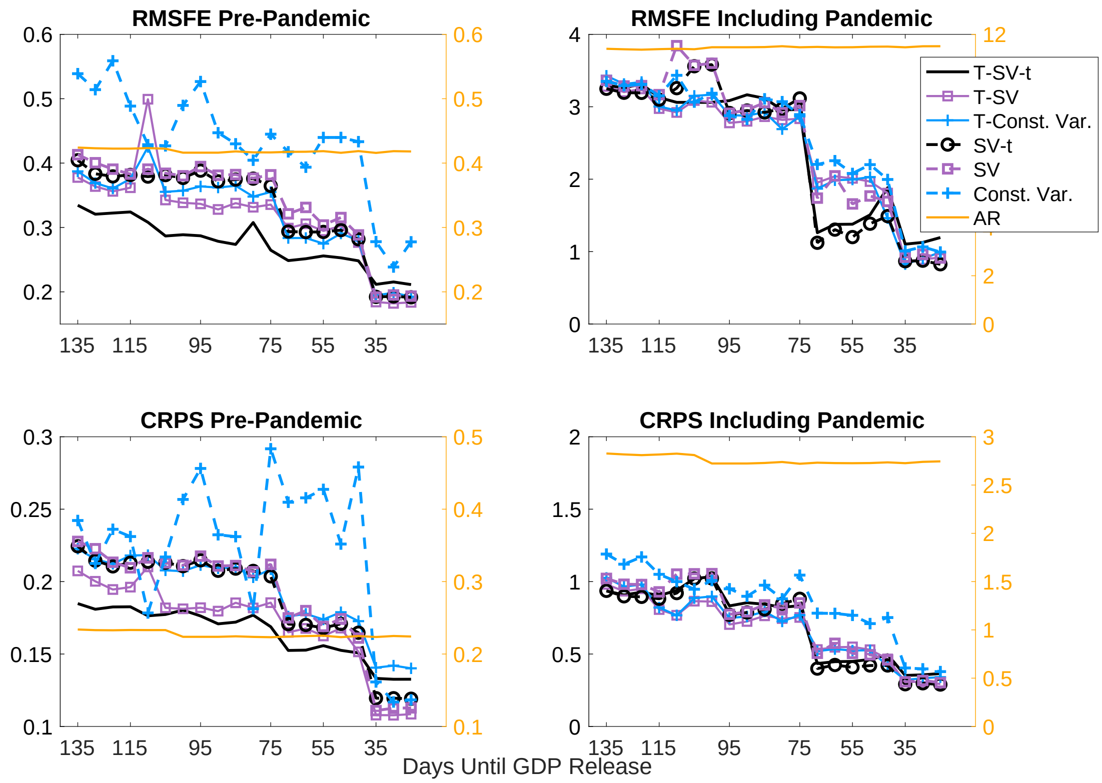


Figure 2: Nowcast performance against alternatives to T-SV-t component.

Notes: Absolute root mean square forecast errors (RMSFE) and continuous rank probability scores (CRPS) over nowcast periods (days ahead of GDP release). Left y-axis: baseline T-SV-t (black solid), T-SV (purple, square marker), T-Const Var (blue, plus), SV-t (dashed black, circle), SV (dashed purple, square), Const Var (dashed blue, plus). Right y-axis: AR(2) (yellow). All models use priors as outlined in section 2.2.

but more so for models that feature stochastic volatility and t-distributed errors. Interestingly, adding the time-varying trend makes less of a difference in the full evaluation sample, and the full model loses out somewhat shortly before GDP release. The simple model without trend and with constant variance fares comparatively well early on in the data release cycle, but then loses out against the other models, particularly in terms of density nowcasts, its relative performance improves compared to pre-pandemic. Overall, this suggests that the most important time-varying feature when including the pandemic is the account for outliers, since it helps the model to identify the Covid-19 pandemic related downturn as temporary. On the other hand, the time-varying trend is less helpful during the pandemic, as models might over-fit the large shock into the trend if outliers are not accounted for.

To understand these results further, we look at the posterior trend and volatility estimates from the T-SV-t BMIDAS with GIGG prior. Figure 3 shows the posterior estimates of the cyclical and trend components (shown separately for pre-pandemic period and Covid-19 period for readability), and the stochastic volatility components of GDP growth and trend (lower panels). The trend-cycle decomposition is intuitive. The cyclical component captures high frequency movements in GDP growth and tracks actual GDP growth (black dashed-dotted lines) well, including over the Covid-19 pandemic, where the cyclical component captures the bulk of the 20% drop in GDP growth and most of the recovery. On the other hand, the trend captures low frequency changes in UK GDP growth, with a gradual slowdown in trend growth since the early 2000s and a temporary trend decrease during the Great Financial Crisis. Throughout the pandemic, the trend remains largely unchanged. Further, the t-distributed volatility estimate of the observation equation shows a sharp and strong increase during the pandemic, by far exceeding the increase observed during the GFC. Hence, the model interprets the extreme movements in GDP growth as transitory in nature and related to a sharp rise in variance in GDP growth rather than in the long-run trend.

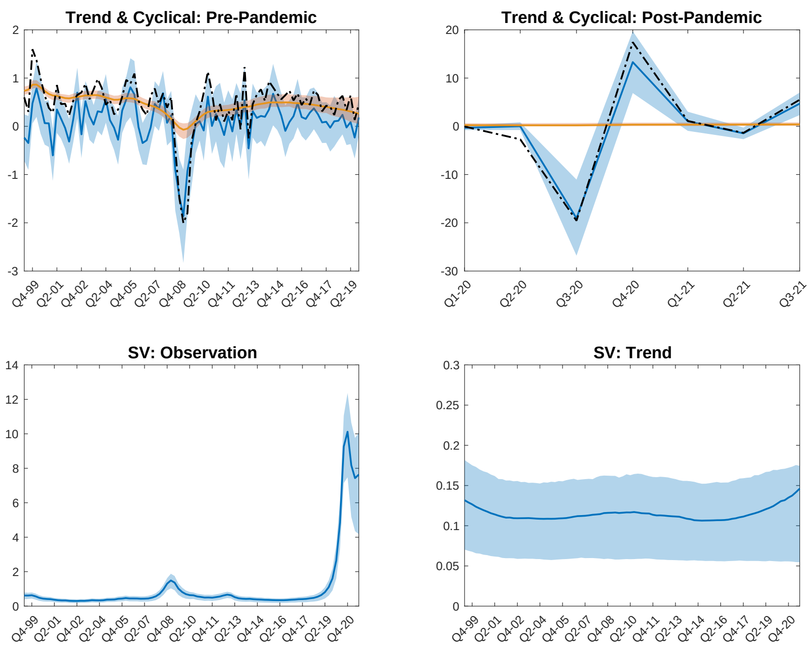


Figure 3: Posterior estimates for trend, cyclical component and stochastic volatilities.

Notes: Full sample, final nowcast period, from T-SV-t BMIDAS, with GIGG prior. Orange: posterior means for trend in GDP growth. Blue: posterior medians of the cyclical component (upper panel) and stochastic volatilities (lower panel). Black dashed line: UK GDP growth. Shaded areas: 95% credible intervals.

Separately identifying slow-moving trends from cyclical fluctuations in semi-structural models is generally challenging and can be sensitive to model choice, sample period and specification. The T-SV-t-BMIDAS is not free from this caveat, and therefore the economic interpretation that we can assign to the trend and cycle estimates can only be suggestive. This said, these flexible model features, in particularly stochastic volatility with t-distributed errors, appear to help achieving a more meaningful and stable trend-cycle identification across nowcast periods. Figure 4 shows the posterior trend and cyclical component from the baseline model compared to the alternative models, over the estimation period until 2019Q4 based on the information set at the first nowcast period, and the 18th nowcast period. The trend and cycle posterior estimates from the Trend-SV-t model remain similar across both nowcast periods, and reflect a slow-moving trend and a cyclical component that captures fluctuations around a positive mean. By contrast, the model with SV but without t-distributed errors identifies a smooth cycle but volatile trend at the early nowcast period, which then shifts around at nowcast period 18. The model with constant variance identifies almost constant trends for both nowcast periods, but with huge uncertainty around them. For nowcast period 18, this model suggest a strongly negative almost fixed trend or intercept in GDP growth and cyclical fluctuations around strongly positive growth. These differences in the trend-cycle identification might explain the relatively weaker nowcast performance of these alternative models, and they suggest the model with SV and t-distributed errors can be preferable also in tranquil times prior to the pandemic.[17]

4.2 Role of group-shrinkage prior with ex-post sparsification

Next, we look at the role of using the GIGG prior and the sparsification step for regularizing the MIDAS regression, by comparing the nowcast results to similar models using alternative priors proposed in the literature. The alternative priors on MIDAS coefficients are

- *GIGG prior without sparsification step.*
- *Horseshoe prior (HS)*: model 1-3 with the horseshoe prior, thus another flexible shrinkage prior, but without group-shrinkage. We derive inclusion probabilities on a group-level using our group-sparsification algorithm from section 2.4.
- *Spike-and-slab prior (SSVS)*: This prior follows George and McCulloch (1993) with a uniform prior on lag-level inclusion probability. Selection of the prior is at the lag-level. We derive

[17] The fat-tailedness of error distributions in the stochastic volatility process proves to be an inherent model feature, as suggested by the posterior distribution of the degrees of freedom parameter showing large mass around small values, see appendix. The appendix also show the trend-cycle identification against alternative priors for the time-varying components and MIDAS coefficients, suggesting a relatively more robust identification for the proposed model.

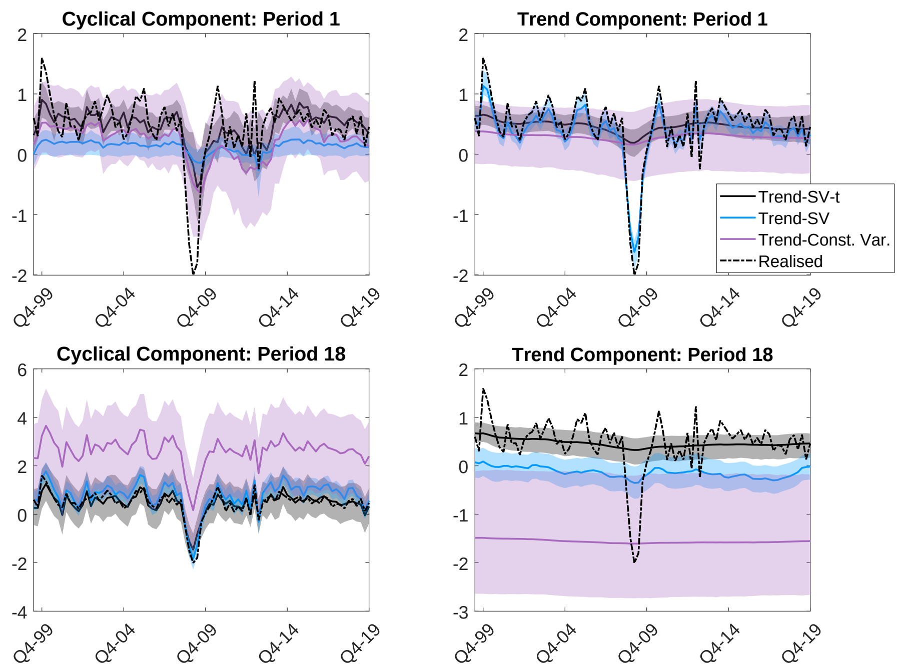


Figure 4: Posterior trend and cycle estimates across specifications.

Notes: Posterior means of trend estimated up until 2019Q4 from the T-SV-t BMIDAS model (black), the T-SV (blue), and T-Const.Var (purple). Estimation at the 1st and 18th nowcast periods. All models estimated with GIGG prior on the MIDAS coefficients.

inclusion probabilities on a group-level using our group-sparsification algorithm.

- *Adaptive group-lasso with spike and slab prior (GAL)*, for Almon lag transformed data. This prior follows Mogliani and Simoni (2021). It is a group-wise spike-and-slab prior, where the slab distribution follows a multivariate Laplace distribution. Hyper-parameters that govern group-inclusion probability are found via an EM algorithm.

First, we compare these prior specifications across models including the time-varying trend with stochastic volatility and t-distributed errors. Figure 6 shows the point and density nowcast evaluation, and Figure 5 visualizes the nowcasts over time for the first and 13th nowcast periods. Exerting group-level shrinkage and taking into account the high-frequency correlation structure via the GIGG prior has preferable performance compared to the horseshoe prior and SSVS prior. Prior to the pandemic, the model with GIGG prior has lower point nowcast errors throughout and almost always better density fit; its nowcasts are somewhat less volatile and show smaller uncertainty bands compared to the alternative priors. When including the pandemic quarters, it is strongest for the first nowcast periods and again starting from 75 days ahead of GDP publication (nowcast period 13 when "hard" indicators are released). All models initially miss the large unprecedented trough.

However, in the first nowcast period, the proposed model with GIGG prior is the only model to indicate the large rebound in activity for Q3-2020. And once real activity indicators for the first reference month are released in nowcast period 13, the model with GIGG prior captures the downward adjustment for Q2-2020 most closely, but it also reflects the highly uncertain environment via the largest uncertainty bands. Nowcast uncertainty then decreases substantially after Q3-2020 for later nowcast periods, albeit it remains a bit larger than for the other models. Hence, with more "hard" macroeconomic information, the model indicates a return toward reduced uncertainty.

Finally, is it the addition of time-varying coefficients or the reliance on group-shrinkage that makes most of the difference for nowcast performance? Or does the group-shrinkage prior work particularly well when combined with the time-varying coefficients? Figure 7 shows results for specifications without trend and constant variance and for alternative priors on the MIDAS coefficients. Here, we also add a specification using the GAL prior, as proposed by Mogliani and Simoni (2021). We see that without time-varying components, the GIGG prior performs rather similarly to the horseshoe and GAL priors, and for density forecasts and when including the pandemic it loses some of its gains seen from Figure 6. The SSVS prior has a weaker and volatile performance without time-variation included, particularly prior to the pandemic. Including the pandemic, the models without time-varying components both with the GIGG prior or the GAL prior achieve some improvement around nowcast period 13 when "hard" data are released, but much less so than the model with time-varying components and GIGG. At nowcast period 18, the models with GIGG prior with and without time-variation perform equally well.

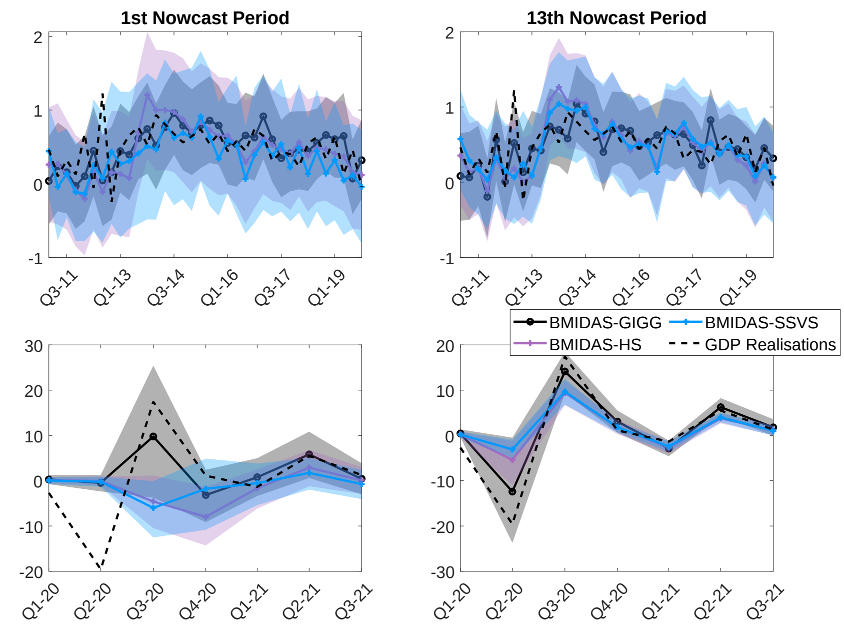


Figure 5: Posterior mean and density nowcasts, for alternatives priors on MIDAS component.

Notes: Nowcasts over period 2011Q1 to 2019Q4 (x-axis) for the 1st and 13th nowcast periods, pandemic quarters (lower panel). Models are as in Figure 6. Shaded areas show 95% credible intervals. Black dashed lines show quarterly GDP growth realizations.

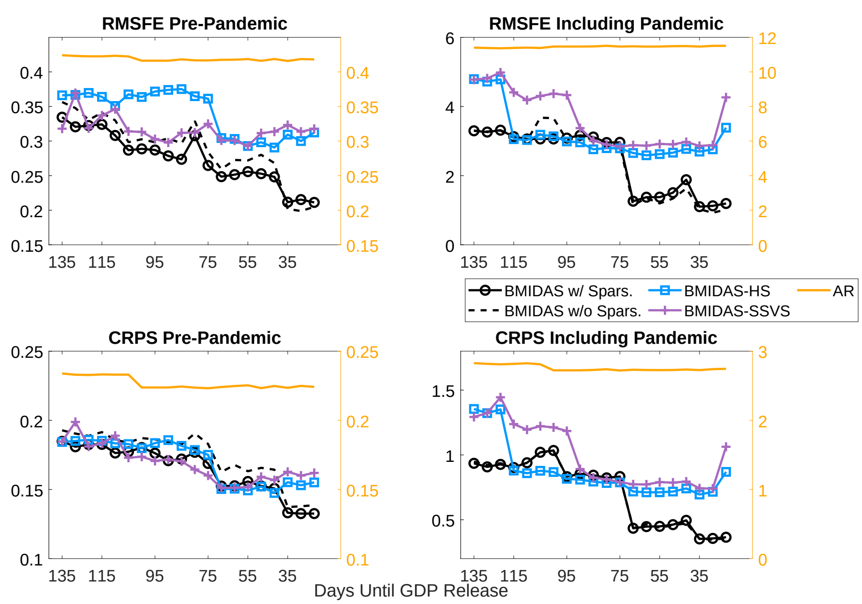


Figure 6: Nowcast performance across alternative priors, T-SV-t-BMIDAS.

Notes: Absolute RMSFE and CRPS for T-SV-t-BMIDAS models with GIGG prior and ex-post sparsification (black solid, circle marker), GIGG without sparsification (black dashed), Horseshoe prior HS (blue, square), Spike-and-slab prior SS (purple, plus). Right y-axis: AR(2) (yellow). All with time-varying trend, SV, t-distributed errors. Models with the GAL prior are only run without time-varying coefficients since the full model with that prior becomes computationally heavy.

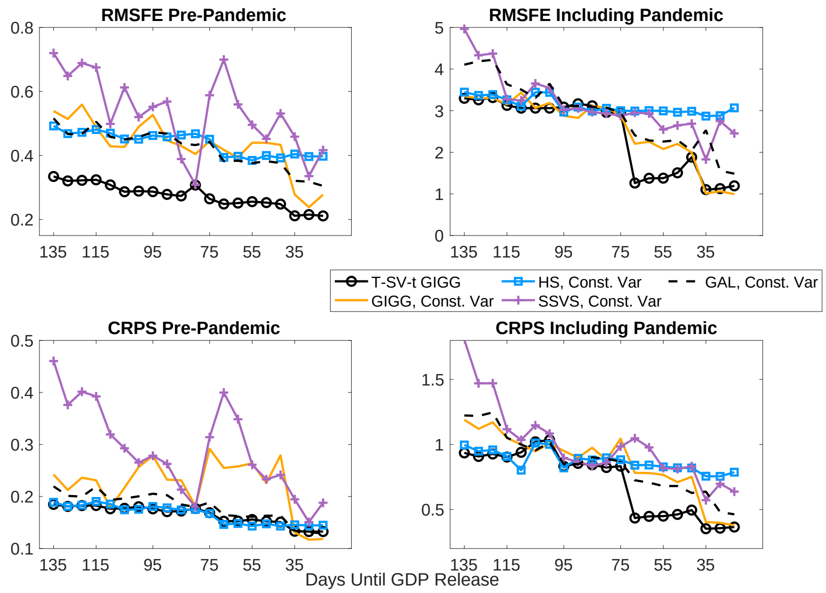


Figure 7: Alternative priors for BMIDAS no trend and constant variance.
Notes: RMSFE and CRPS for T-SV-t BMIDAS with GIGG (solid black line, circle marker), and for BMIDAS models without trend and constant volatility and with different priors on MIDAS coefficients: GIGG (yellow), horseshoe(blue, square marker), SSVS (purple, plus marker), GAL (dashed black).

Overall, the results suggests that the combination of time-varying components and group-shrinkage prior improves nowcast performance over both evaluation periods. As we illustrate in the next section, the GIGG prior shrinks the information set towards a sparse selection of indicators, while still drawing on other indicators to a much lesser extent. And in a model with time-varying coefficients the prior is also able to shift among the most meaningful indicators over the data release cycle and over time, rather then relying on constant signals.

5 Interpreting signals from groups of indicators over time

The sparsification algorithm (9) on the posterior achieves variable selection that allows us to single out which indicators have the largest impact on the predictions of the model over time and over the data release cycle. Figure (8) presents heatmaps for inclusion probabilities for groups of high-frequency lags belonging to each indicator (x-axis) and nowcast period (y-axis), for the pre-pandemic evaluation sample (panel a) and the full evaluation sample (panel b).[18] The upper

[18] Inclusion probabilities are stable over time apart from a pandemic-induced shift, so we focus on averages over sub-samples. The appendix shows the exact inclusion probabilities and standard deviations.

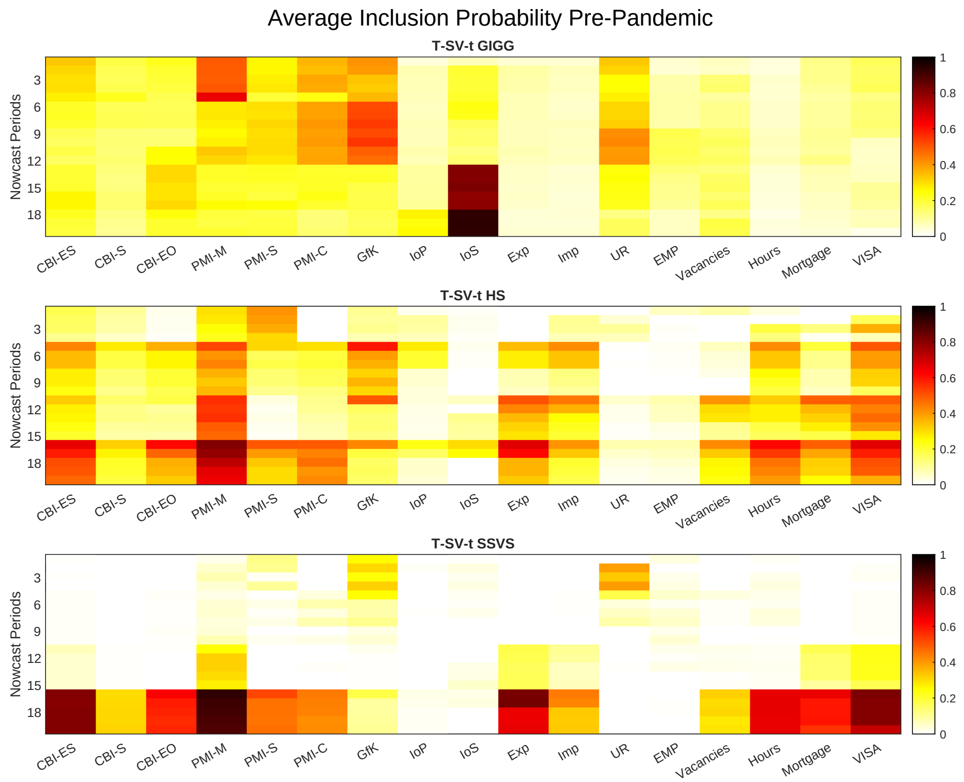
sub-plots in both panels show the T-SV-t BMIDAS with GIGG prior, and the lower sub-plots results from the same model but using the horseshoe prior without group-shrinkage, and the spike-and-slab SSVS prior.

The model with GIGG prior mainly relies on sparse set of indicators at a time, as indicated by only a few red-shaded areas. However, the inclusion probabilities of all other indicators are non-zero as indicated by the yellow colors of different intensity. Thus, information is exploited in line with the idea of "illusion of sparsity" (Giannone et al., 2021), where a broad set of indicators are potentially useful for forecasts, but many signals will be small. Interestingly, the model shifts between groups of indicators over the data release cycle (i.e. moving down the y-axis). Early on, the model exploits a few survey indicators. For the pre-pandemic evaluation sample, these are manufacturing and construction PMIs for very early nowcasts, and then GfK consumer confidence once its observations for the first month of the reference quarter are released. This agrees with earlier findings that survey indicators provide the main early signals for quarterly GDP (Bańbura et al., 2013; Anesti et al., 2017). Later on, labor market data, in particular the unemployment rate, also play a role when released in period 9. Unemployment data are released with a substantial publication lag, as shown in Table 1, which can weaken nowcasting signals. Nonetheless, the model with GIGG prior exploits signals from unemployment more strongly compared to the other priors.

For nowcast period 13, once "hard" economic information get published, we had seen a clear improvement in nowcast performance of the proposed model in Figure 6. From this period on, the model relies almost exclusively on the index of services, an important indicator for the service-oriented UK economy. Instead, the model with horseshoe prior shows a diffuse inclusion pattern across indicators and over nowcast periods, and for both samples. It draws on signals from surveys, real, labor and personal finance indicators, but switches between them over the data release cycle, such as no clear pattern emerges. The SSVS prior, on the other hand, reads little information in early stage of the data release cycle, and only reads intense but broad based signal close to actual GDP release. Overall, by switching from one set of indicators to another, the proposed model with GIGG prior seems to exploit the incoming information more efficiently compared to alternative priors.

With the pandemic included in the evaluation, the higher volatility leads to more intense colors as more signals are being processed, independently of the prior. For the model with GIGG prior, the clear pattern remains of exploiting different signals over the data release cycle, while

a) Pre-Pandemic (2011Q1-2019Q4)



b) Including the Pandemic (2011Q1-2021Q2)

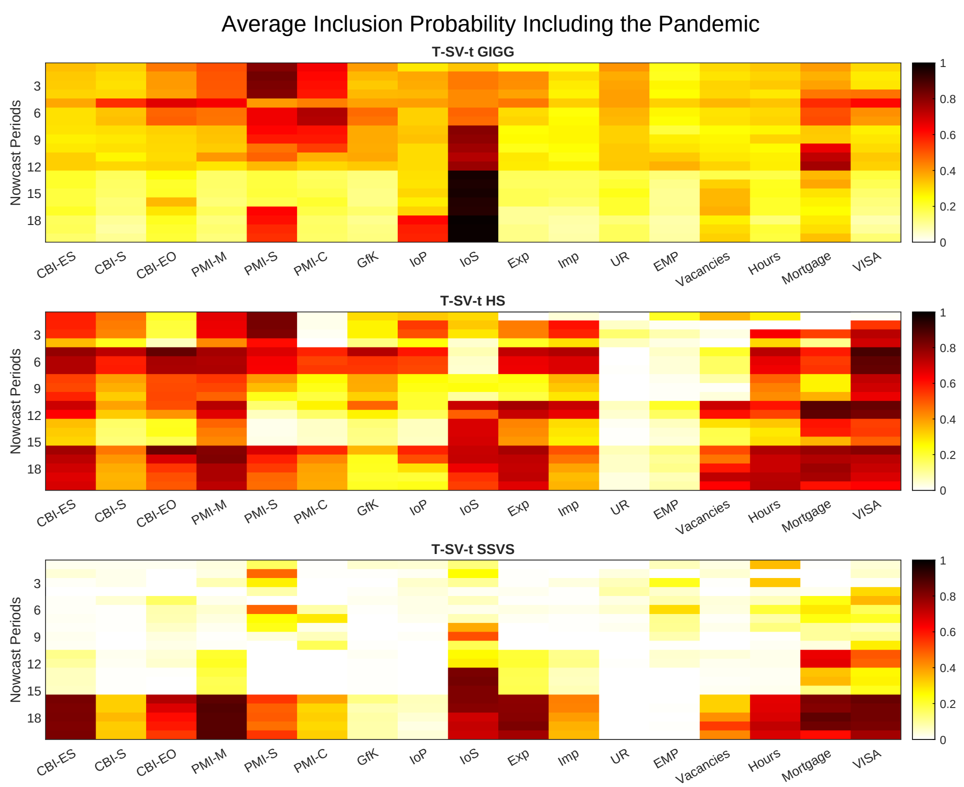


Figure 8: Inclusion probabilities across prior choices, BMIDAS models with trend and SV-t.
Notes: Darker colour indicates higher cumulative posterior inclusion probability for groups of monthly lags of each indicator (x-axis) over nowcast periods (y-axis), from T-SV-t BMIDAS model. Panel a) shows average incl. prob. for evaluation until 2019Q4, panel b) until 2021Q3. Sub-plots show the GIGG prior and horseshoe prior HS, respectively.

still incorporating all other indicators to a much lesser extent. Survey indicators remain most important for early nowcast periods, these are now rather the service and construction PMIs, less so manufacturing. Over nowcast periods 5 to 12, signals from mortgage lending are relevant too, and the index of services turns important, even before its observations for the current quarter get released. Instead, little focus is put on the GFK and labor market data. Thus, the model is able to capture that during the pandemic the economic lockdowns affected mainly the service sector and initially the housing and construction sectors, whereas consumer confidence and manufacturing were less affected and labor market data were distorted by the furlough scheme. In nowcast period 13, there is again a shift towards a strong focus on the index of services, and the inclusion probabilities of most other indicators drop. However, now the model also continues relying on vacancies and mortgage lending, and in nowcast period 18 additional signals stem from the index of production, and again from the service PMI. The horseshoe prior now also relies less on labor market series compared to the pre-pandemic period, and somewhat more on mortgages and VISA consumer spending, but it does not switch away from manufacturing PMIs and other survey indicators. Given that the Covid-19 shock affected specific sectors more than others, a dense solution can represent a disadvantage, as suggested by the weaker nowcast performance of the model with horseshoe prior.

Finally, what is the role of time-varying components for the observed inclusion patterns? Figure B9 in the appendix shows a similar comparison of inclusion probabilities for the models without time-varying trend and with constant variance, with the GIGG prior being compared against the horseshoe and the GAL priors. Overall, the inclusion patterns remain similar. Two notable differences for the model with GIGG prior stand out compared to the case without time-varying components. First, in the model with constant coefficients the GIGG prior is more selective: it selects sub-groups of indicators and the inclusion probabilities of other indicators are close to zero, particularly prior to the pandemic. Thus, while the aggressive group shrinkage helps the model rely on a sub-set of most informative indicators more than on others, the account for the time-varying trend nudges the model to still read weaker signals from all other indicators as well, in line with the "illusion of sparsity" argument. Second, the constant coefficients model does not show a pronounced shift in nowcast period 13 towards a sparser specification; while some probabilities change, the intensity of most colors remains the same. Thus, the time-varying components likely help the model to shift towards signals from "hard" data once these become available. This is not the case for priors without group-shrinkage such as the horseshoe, with or without time-varying components. Overall, this illustrates how efficient group-shrinkage, together with the inclusion of time-varying coefficients, can help select the most meaningful indicators over time and the data release cycle and jointly improve nowcast performance.

6 Conclusion

We have proposed a new Bayesian MIDAS framework, the T-SV-t-BMIDAS model combined with a flexible group-shrinkage prior that accounts for the correlation within groups of indicators. Group-shrinkage and sparsification are achieved in separation. The former serves to regularize the MIDAS coefficients, whereas the latter achieves interpretability and communication of results via a sparsification step on the posterior motivated by Bayesian decision theory.

In an application of the model to nowcasting UK GDP growth, our model is able to capture sharp changes in GDP growth as seen during the Covid-19 pandemic in a relatively timely manner, and works well also in more tranquil times, particularly at a later stage of the data release cycle when "hard" economic indicators are released. The aggressive group-shrinkage helps selecting the most relevant indicators and works particularly well when combined with time-varying components. When accounting for a time-varying trend in GDP growth, the prior achieves group-sparsity by selecting the most relevant indicators, but it also reads a wider range of much smaller signals from all other indicators, in line with the "illusion of sparsity" argument. Allowing for t-distributed errors proves particularly relevant once the Covid-19 pandemic is included. It helps the prior to shift between groups of indicators over time and over the data release cycle, which can important in presence of a large heterogeneous shock.

The proposed model is flexible, combining features and nesting various models as special cases. The derived inclusion probabilities make results more interpretable and easier to communicate over time and over the data release cycle. All this makes an attractive tool not only for nowcasting GDP growth, but also for or a wider range of forecasting applications. The group-shrinkage prior has potential to efficiently regularize information towards a group-sparse set of signals for a range of other data sets where groupings might be present, such as across sectors or countries. We leave the exploration of such alternative applications for future research.

References

Almon, S. (1965). The distributed lag between capital appropriations and expenditures. *Econometrica: Journal of the Econometric Society*, 178–196.

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

Anesti, N., A. B. Galvão, and S. Miranda-Agrippino (2018). Uncertain Kingdom: nowcasting GDP and its revisions. Bank of England Working Paper No 764.

Anesti, N., S. Hayes, A. Moreira, and J. Tasker (2017). Peering into the present: the Bank's approach to GDP nowcasting. *Bank of England Quarterly Bulletin*, Q2.

Angelini, E., G. Camba-Mendez, D. Giannone, L. Reichlin, and G. Rünstler (2011). Short-term forecasts of euro area gdp growth.

Antolin-Diaz, J., T. Drechsel, and I. Petrella (2017). Tracking the slowdown in long-run GDP growth. *Review of Economics and Statistics 99*(2), 343–356.

Antolin-Diaz, J., T. Drechsel, and I. Petrella (2021). Advances in nowcasting economic activity: Secular trends, large shocks and new data. CEPR Discussion Paper No. DP15926.

Armagan, A., D. B. Dunson, and J. Lee (2013). Generalized double Pareto shrinkage. *Statistica Sinica 23*(1), 119.

Babii, A., R. T. Ball, E. Ghysels, and J. Striaukas (2022, July). Machine learning panel data regressions with heavy-tailed dependent data: Theory and application. *Journal of Econometrics*.

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

Bai, J., E. Ghysels, and J. H. Wright (2013). State space models and midas regressions. *Econometric Reviews 32*(7), 779–813.

Bańbura, M., D. Giannone, M. Modugno, and L. Reichlin (2013). Now-casting and the real-time data flow. In *Handbook of economic forecasting*, Volume 2, pp. 195–237. Elsevier.

Bańbura, M. and M. Modugno (2014). Maximum likelihood estimation of factor models on datasets with arbitrary pattern of missing data. *Journal of Applied Econometrics 29*(1), 133–160.

Bank of England, . (2020). How are we monitoring the economy during the Covid-19 pandemic? Bank Overground, https://www.bankofengland.co.uk/bank-overground/2020/how-are-we-monitoring-the-economy-during-the-covid-19-pandemic.

Barbieri, M. M. and J. O. Berger (2004). Optimal predictive model selection. *The annals of statistics 32*(3), 870–897.

Barbieri, M. M., J. O. Berger, E. I. George, and V. Ročková (2021). The median probability model and correlated variables. *Bayesian Analysis 16*(4), 1085–1112.

Baumeister, C., D. Leiva-León, and E. R. Sims (2021). Tracking weekly state-level economic conditions. Technical report, National Bureau of Economic Research.

Berger, T., G. Everaert, and H. Vierke (2016). Testing for time variation in an unobserved components model for the us economy. *Journal of Economic Dynamics and Control 69*, 179–208.

Bhattacharya, A., A. Chakraborty, and B. K. Mallick (2016). Fast sampling with gaussian scale mixture priors in high-dimensional regression. *Biometrika 103*(4), 985–991.

Boss, J., J. Datta, X. Wang, S. K. Park, J. Kang, and B. Mukherjee (2021). Group Inverse-Gamma Gamma Shrinkage for Sparse Regression with Block-Correlated Predictors. *arXiv preprint arXiv:2102.10670*.

Breheny, P. and J. Huang (2015). Group descent algorithms for nonconvex penalized linear and logistic regression models with grouped predictors. *Statistics and Computing 25*(2), 173–187.

Carriero, A., T. E. Clark, and M. Marcellino (2015). Realtime nowcasting with a Bayesian mixed frequency model with stochastic volatility. *Journal of the Royal Statistical Society. Series A,(Statistics in Society) 178*(4), 837.

Carriero, A., T. E. Clark, and M. G. Marcellino (2016). Large vector autoregressions with stochastic volatility and flexible priors.

Carriero, A., T. E. Clark, M. G. Marcellino, and E. Mertens (2021). Addressing COVID-19 outliers in BVARs with stochastic volatility. CEPR Discussion Paper No. DP15964.

Carter, C. K. and R. Kohn (1994). On Gibbs sampling for state space models. *Biometrika 81*(3), 541–553.

Carvalho, C. M., N. G. Polson, and J. G. Scott (2010). The horseshoe estimator for sparse signals. *Biometrika 97*(2), 465–480.

Casella, G., M. Ghosh, J. Gill, and M. Kyung (2010). Penalized regression, standard errors, and Bayesian lassos. *Bayesian Analysis 5*(2), 369–411.

Chakraborty, A., A. Bhattacharya, and B. K. Mallick (2020). Bayesian sparse multiple regression for simultaneous rank reduction and variable selection. *Biometrika 107*(1), 205–221.

Chan, J. C. and A. L. Grant (2016). Modeling energy price dynamics: Garch versus stochastic volatility. *Energy Economics 54*, 182–189.

Chan, J. C. and I. Jeliazkov (2009). Efficient simulation and integrated likelihood estimation in state space models. *International Journal of Mathematical Modelling and Numerical Optimisation 1*(1-2), 101–120.

Chiu, C.-W. J., H. Mumtaz, and G. Pinter (2017). Forecasting with var models: Fat tails and stochastic volatility. *International Journal of Forecasting 33*(4), 1124–1143.

Clark, T. E. (2011). Real-time density forecasts from bayesian vector autoregressions with stochastic volatility. *Journal of Business & Economic Statistics 29*(3), 327–341.

Clark, T. E. and F. Ravazzolo (2015). Macroeconomic forecasting performance under alternative specifications of time-varying volatility. *Journal of Applied Econometrics 30*(4), 551–575.

Cogley, T., S. Morozov, and T. J. Sargent (2005). Bayesian fan charts for UK inflation: Forecasting and sources of uncertainty in an evolving monetary system. *Journal of Economic Dynamics and Control 29*(11), 1893–1925.

D'Agostino, A., L. Gambetti, and D. Giannone (2013). Macroeconomic forecasting and structural change. *Journal of applied econometrics 28*(1), 82–101.

Diebold, F. X., T. A. Gunther, and A. S. Tay (1998). Evaluating density forecasts with applications to financial risk management. *International Economic Review 39*(4), 863–883.

Ferrara, L., M. Mogliani, and J.-G. Sahuc (2022). High-frequency monitoring of growth at risk. *International Journal of Forecasting 38*(2), 582–595.

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

Foroni, C., M. Marcellino, and C. Schumacher (2015). Unrestricted mixed data sampling (MIDAS): MIDAS regressions with unrestricted lag polynomials. *Journal of the Royal Statistical Society: Series A (Statistics in Society) 178*(1), 57–82.

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

Frühwirth-Schnatter, S. and H. Wagner (2010). Stochastic model specification search for gaussian and partial non-gaussian state space models. *Journal of Econometrics 154*(1), 85–100.

George, E. I. and R. E. McCulloch (1993). Variable selection via Gibbs sampling. *Journal of the American Statistical Association 88*(423), 881–889.

Ghysels, E., V. Kvedaras, and V. Zemlys-Balevičius (2020). Mixed data sampling (midas) regression models. In *Handbook of Statistics*, Volume 42, pp. 117–153. Elsevier.

Ghysels, E., P. Santa-Clara, and R. Valkanov (2004). The midas touch: Mixed data sampling regression models.

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

Giannone, D., M. Lenza, and G. E. Primiceri (2019). Priors for the long run. *Journal of the American Statistical Association 114*(526), 565–580.

Giannone, D., M. Lenza, and G. E. Primiceri (2021). Economic predictions with big data: The illusion of sparsity. *Econometrica 89*(5), 2409–2437.

Gneiting, T. and A. E. Raftery (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association 102*(477), 359–378.

Griffin, J. E. and P. J. Brown (2012). Structuring shrinkage: some correlated priors for regression. *Biometrika 99*(2), 481–487.

Guérin, P. and M. Marcellino (2013). Markov-switching midas models. *Journal of Business & Economic Statistics 31*(1), 45–56.

Hahn, P. R. and C. M. Carvalho (2015). Decoupling shrinkage and selection in bayesian linear models: a posterior summary perspective. *Journal of the American Statistical Association 110*(509), 435–448.

Harvey, A. C., T. M. Trimbur, and H. K. Van Dijk (2007). Trends and cycles in economic time series: A bayesian approach. *Journal of Econometrics 140*(2), 618–649.

Hauzenberger, N., F. Huber, and L. Onorante (2021). Combining shrinkage and sparsity in conjugate vector autoregressive models. *Journal of Applied Econometrics 36*(3), 304–327.

Huber, F., G. Koop, and L. Onorante (2021). Inducing sparsity and shrinkage in time-varying parameter models. *Journal of Business & Economic Statistics 39*(3), 669–683.

Huber, F., G. Koop, L. Onorante, M. Pfarrhofer, and J. Schreiner (2020). Nowcasting in a pandemic using non-parametric mixed frequency VARs. *Journal of Econometrics*.

Jacquier, E., N. G. Polson, and P. E. Rossi (2004). Bayesian analysis of stochastic volatility models with fat-tails and correlated errors. *Journal of Econometrics 122*(1), 185–212.

Kapetanios, G., F. Papailias, et al. (2022). Real Time Indicators During the COVID-19 Pandemic Individual Predictors & Selection. Economic Statistics Centre of Excellence (ESCoE).

Kim, C.-J. and C. R. Nelson (1999). Has the US economy become more stable? A Bayesian approach based on a Markov-switching model of the business cycle. *Review of Economics and Statistics 81*(4), 608–616.

Kim, S., N. Shephard, and S. Chib (1998). Stochastic volatility: likelihood inference and comparison with ARCH models. *The Review of Economic Studies 65*(3), 361–393.

Kohns, D. and A. Bhattacharjee (2022). Nowcasting growth using google trends data: A bayesian structural time series model. *International Journal of Forecasting*.

Lenza, M. and G. E. Primiceri (2022). How to estimate a vector autoregression after March 2020. *Journal of Applied Econometrics*.

Lindley, D. V. (1968). The choice of variables in multiple regression. *Journal of the Royal Statistical Society: Series B (Methodological) 30*(1), 31–53.

Makalic, E. and D. F. Schmidt (2015). A simple sampler for the horseshoe estimator. *IEEE Signal*

Processing Letters 23(1), 179–182.

McConnell, M. M. and G. Perez-Quiros (2000). Output fluctuations in the United States: What has changed since the early 1980's? *American Economic Review 90*(5), 1464–1476.

Mogliani, M. and A. Simoni (2021). Bayesian MIDAS penalized regressions: estimation, selection, and prediction. *Journal of Econometrics 222*(1), 833–860.

Ng, S. (2021). Modeling macroeconomic variations after COVID-19. NBER Working Paper No 29060.

O'Hara, R. B. and M. J. Sillanpää (2009). A review of bayesian variable selection methods: what, how and which. *Bayesian analysis 4*(1), 85–117.

Piironen, J., M. Paasiniemi, A. Vehtari, et al. (2020). Projective inference in high-dimensional problems: Prediction and feature selection. *Electronic Journal of Statistics 14*(1), 2155–2197.

Piironen, J. and A. Vehtari (2017). Comparison of bayesian predictive methods for model selection. *Statistics and Computing 27*(3), 711–735.

Piironen, J., A. Vehtari, et al. (2017). Sparsity information and regularization in the horseshoe and other shrinkage priors. *Electronic Journal of Statistics 11*(2), 5018–5051.

Polson, N. G. and J. G. Scott (2010). Shrink globally, act locally: Sparse Bayesian regularization and prediction. *Bayesian Statistics 9*, 501–538.

Ray, P. and A. Bhattacharya (2018). Signal adaptive variable selector for the horseshoe prior. *arXiv preprint arXiv:1810.09004*.

Scruton, J., M. O'Donnell, and S. Dey-Chowdhury (2018). Introducing a new publication model for GDP. *Office for National Statistics*, 1–14.

Simon, N. and R. Tibshirani (2012). Standardization and the group lasso penalty. *Statistica Sinica 22*(3), 983.

Smith, R. G. and D. E. Giles (1976). *The Almon estimator: Methodology and users' guide*. Reserve Bank of New Zealand.

Stock, J. H. and M. W. Watson (2007). Why has US inflation become harder to forecast? *Journal of Money, Credit and Banking 39*, 3–33.

Stock, J. H. and M. W. Watson (2009). Forecasting in dynamic factor models subject to structural instability. *The Methodology and Practice of Econometrics. A Festschrift in Honour of David F. Hendry 173*, 205.

Stock, J. H. and M. W. Watson (2016). Core inflation and trend inflation. *Review of Economics and Statistics 98*(4), 770–784.

Wang, H. and C. Leng (2008). A note on adaptive group lasso. *Computational statistics & data analysis 52*(12), 5277–5286.

Woody, S., C. M. Carvalho, and J. S. Murray (2021). Model interpretation through lower-dimensional posterior summarization. *Journal of Computational and Graphical Statistics 30*(1), 144–161.

Xu, X. and M. Ghosh (2015). Bayesian variable selection and estimation for group lasso. *Bayesian Analysis 10*(4), 909–936.

Xu, Z., D. F. Schmidt, E. Makalic, G. Qian, and J. L. Hopper (2016). Bayesian grouped horseshoe regression with application to additive models. In *Australasian Joint Conference on Artificial Intelligence*, pp. 229–240. Springer.

Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association 101*(476), 1418–1429.