A Flexible Bayesian MIDAS Approach for Interpretable Nowcasting and Forecasting*

David Kohns[†] Galina Potjagailo [‡]

Heriot-Watt University, UK Bank of England

September 22, 2022

Work in progress.

Abstract

We propose a flexible model for nowcasting and forecasting that accounts for slow-moving trends, stochastic volatility and t-distributed errors in the low-frequency target, and exploits high frequency information via Bayesian MIDAS. To regularise the multivariate MIDAS component we employ a global-local prior that flexibly imposes three-tiered shrinkage (overall, between indicators, and within lags of an indicator). We enhance the prior with a sparsification step motivated by Bayesian decision theory which allows to communicate importance of predictors via inclusion probabilities over the data release cycle. We use our model to nowcast GDP growth in the UK over the period 1999 to 2021. Results show that accounting for a time-varying trend and t-distributed stochastic volatility substantially improves nowcast performance. Moreover, the shrinkage prior allows to select a sparse group of the most informative indicators over time, leading to improvements in nowcast performance compared to dense specifications such as a dynamic factor model. When the Covid-19 pandemic sets in, the model switches to reading stronger signals from indicators for services, which reflected spending shifts related to lockdowns, and less from production surveys. This helps to nowcast the initial recovery after the shock, and to update the nowcast for the pandemic-induced trough sooner.

Keywords: Forecasting, Mixed Frequency, Covid-19, Grouped Horseshoe Prior, Decision Analysis.
JEL Codes: C11, C32, C44, C53, E37

** Disclaimer:* The views expressed in this paper are solely those of the authors and do not necessarily represent those of the Bank of England.

[†]Corresponding author. Email: david.kohns94@googlemail.com
[‡]Email: galina.potjagailo@bankofengland.co.uk

1 Introduction

The Covid-19 pandemic has raised new challenges for nowcasting and forecasting. The large size of the shock caused parameter instabilities in time series methods. Moreover, the shock hit economic sectors in a strongly heterogeneous way and timely survey indicators provided only weak signals for economic activity, which further challenged models that rely on co-movement across many such indicators. More generally, this episode has illustrated the need for more flexible model features that account for time-variation and regime shifts, large shocks, and heterogeneous signals. This remains relevant also in aftermath of the pandemic, to account for further large economic distortions such as those induced by Russia's invasion of Ukraine and the high inflation rates observed across advanced economies.

The inclusion of time-varying trends and stochastic volatility into unobserved component (UC), dynamic factor models (DFM) and vector-autorgressive (VAR) models have been shown to be beneficial for forecasting (Stock and Watson, 2009; Clark, 2011; D'Agostino et al., 2013; Berger et al., 2016; Carriero et al., 2016; Antolin-Diaz et al., 2017). Since the pandemic, methods to deal with extreme observations have ranged from down-weighting them via outlier components and t-distributed volatilities (Lenza and Primiceri, 2022; Carriero et al., 2021; Antolin-Diaz et al., 2021) to dropping them altogether (Schorfheide and Song, 2021).[1]

In this paper, we propose the Trend-SV-t-BMIDAS model. It combines three distinct model features that to our knowledge have not been exploited jointly so far: 1) an unobserved component block that accounts for a time-varying trend and stochastic volatility allowing for extreme observations in GDP growth, 2) a Mixed Data Sampling (MIDAS) regression that exploits information from higher frequency indicators and 3) a flexible group-shrinkage prior for regularisation and variable selection. Our contribution is two-fold. First, we propose a general setting for nowcasting and forecasting that accounts for shifts over time in the low-frequency target and that exploits potentially heterogeneous high frequency information via multivariate MIDAS. While our setting nests existing models as special cases when model components are shut-down, we show that combining such features leads to strong nowcasting performance, including in and around the Covid-19 pandemic. Second, for regularising the MIDAS component we employ the GIGG (Group Inverse-Gamma Gamma) prior. This allows for three-tiered shrinkage—overall, between indicators, and within lags of an indicator—informed by the correlation structure in the data. We enhance the prior with a new sparsification algorithm for variable selection that allows to communicate signals from indicators in an intuitive way.

Specifically, the Trend-SV-t-BMIDAS features a time varying trend component (Trend) to

[1] Alternatively, various studies have exploited new data sources at a daily or weekly frequency in factor models (Ng, 2021; Baumeister et al., 2021) or non-parametric predictive regressions (Woloszko, 2020; Huber et al., 2020; Kapetanios et al., 2022).

capture low-frequency changes in GDP growth, and stochastic volatility processes that account for fat tails (SV-t) in the conditional GDP growth distribution and in the trend. These features capture gradual shifts in long-run GDP growth or its variance, and allow for extreme observations that, if un-accounted for, could blur signals from the high-frequency indicators. For these components, we employ normal-type priors on the state variances using non-centred state space methods (Frühwirth-Schnatter and Wagner, 2010).

Conditionally on the trend, cyclical fluctuations in GDP growth are linked directly to a set of higher frequency indicators and their lags via parametric functional constraints in a multivariate MIDAS regression (Ghysels et al., 2007, 2020). This approach is computationally less demanding compared to models that address the frequency mismatch via a state space representation (Bai et al., 2013). And since the MIDAS regression does not rely on co-movement in the data, it can exploit heterogeneous signals from each indicator and its lags, which can be desirable during a large heterogeneous shock such as Covid-19. However, multivariate MIDAS models can be highly parameterised, and therefore prone to overfitting (Frühwirth-Schnatter and Wagner, 2010; Huber et al., 2019), particularly when adding potentially many indicators and time-varying features (Carriero et al., 2015). We impose Almon-lag polynomial restrictions which reduce the number of parameters, but the MIDAS component can suffer from multi-collinearity due to the high degree of serial correlation present in the mixed frequency lags.

The GIGG shrinkage prior takes the grouping between high-frequency lags and the typically large correlation among them into account. It was initially proposed for group-shrinkage in a panel data setting in environmental health studies by Boss et al. (2021), but has not been exploited within a time series context. The prior shrink the impact of indicators in a continuous manner, without shrinking any to zero, which makes communication of signals from indicators challenging. We propose to combine the prior with a new sparsification algorithm motivated by Bayesian decision theory. The algorithm selects high-frequency lag groups that best summarise the predictions of the model in the spirit of Hahn and Carvalho (2015) and Chakraborty et al. (2020) and shrinks others to exact zeros. This allows to derive inclusion probabilities for individual (or groups of) indicators at each point in time.

In an empirical application, we employ the Trend-SV-t-BMIDAS model with GIGG prior to nowcast quarter-on-quarter GDP growth in the United Kingdom using a set of monthly macroeconomic indicators. We evaluate nowcasts over the data release cycle in a pseudo-real-time setting over the sample period 1999 to 2021, distinguishing between a pre-pandemic sample and the full sample including the Covid-19 shock. We evaluate the model against simpler specifications that do not include time-varying components, where the prior does not impose group shrinkage, and against alternative specification of the MIDAS component.

We present three main empirical results. First, the posterior estimates of the time-varying trend and stochastic volatility components for UK GDP growth conform to economic intuition.

We document a gradual decline in the trend component in UK GDP growth since the early 2000s. Throughout the Covid-19 pandemic period, trend GDP growth and its volatility remain largely unchanged, whereas the cyclical component absorbs most of the shock and the volatility of GDP growth shows a sharp increase. The inclusion of stochastic volatilies with t-distributed errors helps the model to interpret the extreme movements in GDP growth seen during the pandemic quarters as transitory in nature, since it helps to discount noisy observations which otherwise would be interpreted as trend shifts.

Second, the proposed model provides gains against simpler specifications where the time-varying components are shut down or the MIDAS component is modelled differently. Accounting either for a time-varying trend or for SV with t-errors substantially improves nowcasts, and accounting for both provides further gains, particularly prior to the pandemic. Compared to alterntive models where high frequency indicators are exploited through a combination of univariate MIDAS without group-shrinkage, or summarised by a dynamic factor, the proposed multivariate BMIDAS with group shrinkage performs better in late nowcast periods for which real activity indicators for the reference quarter are available. The relative improvement is particularly pronounced once the Covid-19 shock is included in the sample, where our model detects the initial economic trough earlier and nowcasts the economic recovery more precisely.

Third, we unpack the signals that the Trend-SV-t BMIDAS exploits by presenting inclusion probabilities of the indicators. We find that the group-shrinkage through the GIGG prior lets the model draw on a sparse subset of indicators throughout the data release cycle. It reads signals from survey indicators early in the data release cycle, and then shifts towards signals from a few real activity indicators, particularly the index of services. When including the pandemic period into the evaluation, the model concentrates more heavily on service-related surveys and activity indicators, as well as indicators from the housing sector, both of which reflected disruption from lockdowns and consumption shifts during the pandemic. Other models, on the other hand, continue to read dense signals from a wide range of indicators. This finding contributes to the general debate on the "illusion of sparsity" in macroeconomic forecasting (Giannone et al., 2021; Fava and Lopes, 2021). We find that our sparse model outperforms denser alternatives and seems better able to adjust to the heterogeneous nature of the Covid-19 shock. Overall, our analysis suggests that it is important to account for grouping structure across indicators when analyzing sparsity patterns in applications with macroeconomic data.

We provide sensitivity analysis which shows that our model retains much of its competitive performance when adopting a U-MIDAS transformation, although the Almon lag restrictions help later in the data release cycle. On the other hand, we find that holding MIDAS coefficients fixed at their pre-pandemic levels, in the spirit of Schorfheide and Song (2021), markedly decreases nowcast precision. This indicates that the model coefficients adjust flexibly once data signals come in during the pandemic quarters, which helps capturing the large shock.

The application to nowcasting highlights the ability of the model to exploit timely signals—which has been of particular interest for predictions during the Covid-19 pandemic, when economic conditions deteriorated very fast and the typically heavily relied upon survey indicators temporarily lost much of their association with GDP growth. However, the model can be applied to a range of applications. The inclusion of time-varying trends makes it suitable for medium-term GDP or inflation forecasting, and group-wise shrinkage can be useful for regularising a range of disaggregated time series data.

Our analysis relates to the literature on multivariate MIDAS models with non-linear components. 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 accounting for non-linear features in the target variable via a time-varying trend and stochastic volatilities that account for extreme observations, as these features have been found to improve predictive performance in a range of other models.

Our shrinkage approach relates to the literature on global-local priors. These priors are characterised by two heavy- tailed scale processes that simultaneously shrink globally and on individual covariate level (Polson and Scott, 2010; Polson et al., 2014), where the horseshoe prior (Carvalho et al., 2010; Bhadra et al., 2019) is particularly popular. Conventional global-local shrinkage priors, however, do not feature group-shrinkage and can suffer from random covariate selection and bad mixing for highly correlated designs (Boss et al., 2021; Piironen et al., 2020). Carriero et al. (2015) apply shrinkage on individual level with implicit grouping via a Minnesota type prior instead, which imposes a deterministic degree of penalisation that increases with the lag-length. We instead employ the GIGG prior that imposes adaptive group-shrinkage accounting for the grouping and correlation of lags.

Our setting nests various existing models from the literature as special cases. When shutting down the Trend-SV-t component and setting the hyper-parameters that impose group-wise and within-group shrinkage to zero, we get a multivariate Bayesian MIDAS model with a horseshoe prior, similar to Kohns and Bhattacharjee (2020). When allowing only for stochastic volatility, the model is closest to Carriero et al. (2015). When shutting down the Trend and SV-t component but imposing group-shrinkage, the resulting model is closest to Mogliani and Simoni (2021) who propose a multivariate BMIDAS model with a group-shrinkage prior that has a spike-and-slab structure. Finally, when modelling the cyclical fluctuations via a DFM rather than as MIDAS structure, the model comes close to Antolin-Diaz et al. (2021).

The remainder of the paper is structured as follows. Section 2 presents the Trend-SV-t Bayesian MIDAS, the GIGG prior and the proposed sparsification step. Section 3 outlines the data set and setup of our application. Section 4 present the results on the in-sample model features, nowcast performance and variable inclusion probabilities. Section 5 concludes. Details on the methodology and additional results can be found in appendices A and B.

2 Trend-SV-t Bayesian MIDAS with flexible priors

In this section, we discuss our methodological contributions to the BMIDAS framework. Firstly, we outline the main features of the model. Then, we present the GIGG group-shrinkage prior which we argue is well suited to regularise the multivariate MIDAS component due to its ability to take the mixed-frequency grouping and correlation structure into account. Lastly, we detail the sparsification algorithm that allows to communicate which signals the model reads from high-frequency variable groups.

2.1 The Trend-SV-t-BMIDAS model

The proposed Trend-SV-t BMIDAS model is a flexible unobserved components model that includes three main features: (1) a time-varying trend component that captures latent slow-moving changes in GDP growth, (2) a multivariate MIDAS regression component that captures cyclical macroeconomic information from a range of higher-frequency (monthly) indicators, and (3) fat-tailed stochastic volatility processes to model error clustering in GDP growth and in the trend, where leptokurtic tails discount potential outliers.
The model takes the following state-space form:

$$\begin{aligned} y_t &= \tau_t + \theta' Z_{t-h}^{(m)} + \sqrt{\lambda_t} e^{\frac{1}{2}(h_0 + w_h \tilde{h}_t)} \tilde{\epsilon}_t^y, \\ \tilde{\epsilon}_t^y &\sim N(0,1),\ \lambda_t \sim IG(\nu/2, \nu/2) \end{aligned} \tag{1}$$

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

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

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

The trend component τ_t follows a driftless random walk that captures low frequency changes in GDP growth. This reflects slow moving shifts in general economic conditions which would not be captured by high-frequency information alone (Kim and Nelson, 1999; McConnell and Perez-Quiros, 2000; Antolin-Diaz et al., 2017).[2]

Conditionally on the trend, the component $Z_t^{(m)}$ captures cyclical fluctuations in GDP growth. $Z_t^{(m)} = (z_{1,t}^{(m)}, \cdots, z_{K,t}^{(m)})'$ is observed for each of the K high-frequency indicators at m, potentially non equidistant, intervals between $t-1$ and t. In our application, we consider macroeconomic indicators which are observed at monthly frequency, so that $m = 3$. The pa-

[2] A different non-stochastic approach to disciplining models to account for long-run growth features is presented in Giannone et al. (2019). Here, the authors enforce iterative forecasts to return to long-run cointegrating equilibria.

rameters $\theta = (\theta_1', \cdots, \theta_K')'$ measure the response to changes in the K macroeconomic indicators, each with $L_k = 6$ monthly observations.[3] θ_k features $p_k + 1$ parameters that act as a linking function between higher frequency and lower frequency observations. The functional form of these parameters further depend on the type of MIDAS employed (see for an overview of MIDAS methods Ghysels and Marcellino (2018)). Due to its parsimony, we use the linear Almon lag polynomial in the spirit of Almon (1965) that has been recently popularised for high dimensional mixed frequency forecasting applications by Mogliani and Simoni (2021).[4] Parsimony is induced in linear Almon-MIDAS by assuming a $p_k << L_k$ polynomial process of the coefficients across high-frequency observations, which can be further improved with economically relevant end-point restrictions (Smith and Giles, 1976). For our application below, we assume 5 monthly lags (s.t $p_k + 1$ spans 6 months in total), a third degree polynomial ($p_k = 3$) as proposed by Mogliani and Simoni (2021) and two end-point restrictions to ensure that the weight profile peters out smoothly to 0 (see Appendix A for details). Although Almon lags partially address over-parametrisation, the MIDAS component can still suffer from multi-collinearity due to the high degree of serial correlation present in the mixed frequency lags $Z_t^{(m)}$. This will be addressed with our shrinkage prior in section 2.2.2.

Finally, we allow for stochastic volatility processes $\tilde{h}_t, \tilde{g}_t$ for GDP growth y_t and the latent trend τ_t, respectively. The volatility processes follow driftless random walks (3), and are non-centred after Frühwirth-Schnatter and Wagner (2010). This exerts stronger shrinkage on the state standard deviations, w_h and w_g so as to control the variability of the state process.[5] The more commonly employed centred SV process h_t, can be exactly recovered by noticing that $h_t = h_0 + w_h \tilde{h}_t$ (likewise for g_t). Despite the stronger shrinkage of the state variation, random walk behaviour might be inappropriate to capture the recent unprecedented size and short-lived nature of the Covid-19 shock. We therefore allow the observation equation to have fat-tails via λ_t which enforces t-distributed errors with ν degrees of freedom. This discounts large contemporaneous movements in y_t and thus limits the propagation of outliers to the posteriors of the model components.

---

[3] Three observations correspond to leads and three are lags in relation to the quarterly time-steps, t.

[4] 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. (2007), but have the advantage of being compatible with off the shelf shrinkage methods.

[5] Since w_g and w_h, directly appear in the observation and state equation, which control the overall state smoothness, one can apply conjugate normal priors which exert stronger shrinkage than inverse-gamma priors. This approach has been used in many large time-varying parameter models (Huber et al., 2019; Chan, 2017b; Koop and Onorante, 2019).

2.2 Bayesian Setup

To regularise parameter estimation variance, we make use of Bayesian priors. We first present the general prior framework that gives rise to the conditional posteriors, before detailing in 2.2.2 the form and intuition of the GIGG prior.

We consider priors of the form:

$$\pi(\zeta) = \pi(\theta)\pi(\boldsymbol{\tau})\pi(\tilde{\boldsymbol{h}})\pi(\tilde{\boldsymbol{g}})\pi(\phi)\pi(\boldsymbol{\lambda}|\nu)\pi(\nu), \tag{4}$$

where ζ stacks all unknown parameters into one vector, bold-faced letters refer to time ordered vectors (e.g., $\boldsymbol{\tau} = (\tau_1, \cdots, \tau_T)'$) and ϕ collects any remaining state parameters $(\tau_0, h_0, g_0, w_h, w_g)'$. The priors, except for those for $\boldsymbol{\lambda}, \nu$, are independent which allows for convenient Gibbs sampling of the conditional posteriors.

2.2.1 Priors for the cyclical component

We consider normal priors for the MIDAS parameters θ:

$$\pi(\theta) \sim N(0_{\sum_{k=1}^{K}(p_k+1)}, \Lambda_*), \tag{5}$$

where a prior mean vector of zero implies shrinkage toward sparsity and the prior variance parameters along the diagonal of Λ_* control the amount of shrinkage toward zero. These will be populated with parameters of the GIGG prior, as discussed in the next sub-section. But the structure of the prior allows for any independent shrinkage prior, see for example Polson and Scott (2010) for an overview of common shrinkage priors. Conditional on the other model parameters, the posterior is normal:

$$\begin{gathered} \theta|\boldsymbol{y}, \bullet \sim N(\overline{\theta}, \overline{\Lambda}_*^{-1}) \\ \overline{\Lambda}_* = (\boldsymbol{Z^{(m)}}'\Lambda_t^{-1}\Lambda_h^{-1}\boldsymbol{Z^{(m)}} + \Lambda_*^{-1}),\ \overline{\theta} = \overline{\Lambda}_*^{-1}(\boldsymbol{Z^{(m)}}'\Lambda_t^{-1}\Lambda_h^{-1}\tilde{\boldsymbol{y}}), \end{gathered} \tag{6}$$

where $\tilde{\boldsymbol{y}} = \mathbf{y} - \boldsymbol{\tau}$, $\Lambda_t = diag(\lambda_1, \cdots, \lambda_T)$, $\Lambda_h = diag(e^{h_1}, \cdots, e^{h_T})$, and $\boldsymbol{Z^{(m)}} = (Z_1^{(m)}, \cdots, Z_T^{(m)})'$. 6 highlights the effects of the stochastic volatility and fat-tail imposed in the model. Both $\tilde{\boldsymbol{h}}$ and $\boldsymbol{\lambda}$ act to effectively discount variation in the covariance matrix as well as the fit with $\tilde{\boldsymbol{y}}$. The scales in Λ_t move inversely with irregular shocks in conditional GDP growth, while Λ_h account for slow moving error variance.

Due to the availability of known conditional distributions, we make use of Gibbs sampling to draw inference on the parameters (see A.2), where we use 5000 draws as burn-in and retain 5000 for inference. *** PLEASE ADJUST:*** In situations where Λ_* is very high-dimensional, particularly when $\sum_{k=1}^{K} p_k + 1 >> T$, we make use of the fast sampling algorithm of Bhattacharya et al. (2016) which reduces computational complexity for the regression parameter sampling step from $\mathcal{O}((\sum_{k=1}^{K} p_k + 1)^3)$ (Cholesky based sampling algorithms) to $\mathcal{O}(T^2 \times (\sum_{k=1}^{K} p_k + 1))$.[6]

2.2.2 Group-shrinkage prior (GIGG) on the multivariate MIDAS structure

The multivariate MIDAS framework is highly parameterised since it involves K indicators, and for each indicator k, a group of L_k lower frequency (quarterly) vectors are created corresponding to each available observation at higher (monthly) frequency. This gives two potentially interdependent dimensions relevant for efficient shrinkage. First, not all of the K groups are equally relevant for prediction, which calls for adaptive *shrinkage across groups.* Second, the fact that consecutive higher frequency observations enter the model via the implied data transformations induces correlation *within the lag group.* Such correlation can be high for time-series applications, even with U-MIDAS sampled data (Ghysels et al., 2007), and more so when Almon lag polynomials impose a structure on the lags. This can cause mixing issues with aggressive shrinkage priors which typically seek to shrink individual covariates aggressively to zero or not at all. Finally, the two dimensions can be interdependent, since the degree of correlation across lags can matter for the relative impact that the lag group has for predicting the target.

We therefore use a three-tiered shrinkage prior that addresses these issues jointly, as it adaptively shrinks groups to zero, and simultaneously accounts for the degree of correlation within the group. The prior, named the group inverse-Gamma Gamma (GIGG) prior (Boss et al., 2021), is specified as

$$\begin{aligned} \theta_{k,j} &\sim N(0, \vartheta^2 \gamma_k^2 \varphi_{k,j}^2), \quad \forall j \in \{0, \cdots, p_k + 1\} \\ \vartheta \sim C_+(0,1), \quad \gamma_k^2 | a_k &\sim G(a_k, 1), \quad , \varphi_{k,j}^2 \sim IG(b_k, 1), \end{aligned} \tag{7}$$

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.

The parameters ϑ, γ_k and $\varphi_{k,j}$ govern the three-tiered shrinkage. While ϑ^2 controls the overall level of sparsity, γ_k^2 acts as a shrinkage factor that enables pushing the impact of group k jointly close to zero, and $\varphi_{k,j}^2$ controls how correlated group member j is within k. The hyper-parameters a_k and b_k summarise our prior guess on the relative importance of group-level shrinkage relative to shrinkage of within-group correlation (see Appendix A for a visualisation of the a-priori behaviour). The lower a_k is set relative to b_k, the stronger the group-level shrinkage, and the larger the prior correlation among the group's regression parameters. Such a choice favors group-sparse posteriors, where groups are shrunk jointly to zero, but the relevant groups

[6] Details of the sampling algorithm can be found in A.1

feature a high degree of correlation among individual lags, instead of a heterogeneous shrinkage within group. This can be an appropriate choice for models that use Almon transformed lags, and even in unrestricted MIDAS applications where lags are highly correlated, as suggested by Boss et al. (2021). In our application, we follow this intuition and we set $a_k = 1/T, b_k = 0.5\ \forall k$. In appendix B.3 we report nowcast results for different hyper-parameter choices. In situations in which prior knowledge exists that only selected lags are important, a relatively small b_k helps to shrink individual lags within a group.

The GIGG prior belongs to the general global-local prior framework, in which two heavy-tailed scale processes simultaneously shrink globally and on individual covariate level (see Polson and Scott (2010) and Polson et al. (2014)). Among these, the horseshoe prior of Carvalho et al. (2010) is particularly popular due to its excellent empirical and theoretical properties (Bhadra et al., 2019; Van Der Pas et al., 2014). Conventional global-local shrinkage priors such as the horseshoe, however, do not feature group shrinkage via γ_k^2 and can suffer from random covariate selection and bad mixing for highly correlated designs (Boss et al., 2021; Piironen et al., 2020; Giannone et al., 2021), which is typically the case in mixed-frequency regressions. Previous BMIDAS priors such as in Carriero et al. (2015) apply shrinkage on individual level with implicit grouping via a Minnesota type prior to address this, but the Minnesota prior does not impose adaptive shrinkage across covariates k and imposes a deterministic level of penalisation that increases with the lag-length. Closest to our prior setup for MIDAS regression is Kohns and Bhattacharjee (2020) who use simple horseshoe prior regularisation. We will use this prior as a benchmark for the group-prior model.

A special trait of the GIGG prior is that it nests the exact group-horseshoe prior when $a_{gk} = b_{gk} = 0.5$, for a group-size of one and for other combinations of the hyper-parameters, follows a correlated normal beta prime distribution akin to Armagan et al. (2013), $\gamma_k^2 \varphi_{k,j}^2 \sim \beta'(a_k, b_k)$.[7] It therefore inherits the characteristics of the β' distribution in that it can allow for aggressive shrinkage due to large prior mass on zero, but also allows groups to be virtually unregulated when the effect on the target is large due to it's heavier than exponential tails, which makes the prior appropriate for group selection tasks (Piironen et al., 2017).[8]. The shrinkage properties of the prior are further elucidated in Appendix (A.1).

In contrast to the proposed GIGG prior, other group-prior implementations such as the group-lasso suggested in the literature (Casella et al., 2010; Xu and Ghosh, 2015), are known for over-regularising signals due to their exponential tails and under-regularising noise due to less mass on zero compared to horseshoe type priors. Xu and Ghosh (2015) remedy the latter

[7] This represents a departure from the group-horseshoe prior of Xu and Ghosh (2015), which does not reduce to a simple horseshoe for a group size of one.

[8] Unregulated groups with large effects on the target is a trait that is shared on an individual covariate level with the horseshoe prior Carvalho et al. (2010)

behaviour of the group-lasso by including group-spike-and-slab variable selection. However, these priors assume a uniform level of shrinkage within group, hence, don't allow for inference on the correlation via a covariate level scale. Further, the spike-and-slab with point-mass on zero is well known to mix poorly in high dimensions and with correlated groups (Ishwaran et al., 2005; Piironen et al., 2017). For these reasons, we do not consider lasso priors, but refer to Boss et al. (2021) for further prior comparisons to the GIGG prior.

2.2.3 Priors for the latent states & fat-tails

The other priors we employ are standard. 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); McCausland et al. (2011) that allow representing the entire conditional state posterior as a tractable normal distribution. This increases sampling efficiency compared to Kalman filter based techniques such as Carter and Kohn (1994); Frühwirth-Schnatter (1994), and allows use of computationally efficient sparse matrix operations (see A.1 for a more detailed exposition). To derive the posteriors of $(\tilde{\boldsymbol{h}}, \tilde{\boldsymbol{g}})$ we use the approximate sampler of Kim et al. (1998). Since latent states are prone to overfitting in heavily parameterised models Frühwirth-Schnatter and Wagner (2010) we put normal priors on the state variances to control their variation. These exert stronger shrinkage than commonly employed inverse-Gamma priors for variance parameters (see Frühwirth-Schnatter and Wagner (2010) for more discussion on state variance priors). The conditional posteriors are further exposed in A.1. Lastly, for the degrees of freedom hyper-parameter ν of the Inverse-Gamma distribution of λ_t we assume a relatively non-informative uniform prior which results in a non-standard distribution detailed in appendix A.1. To draw from that posterior, we make use of a Metropolis-within-Gibbs step.

2.3 Sparsification step for the GIGG prior

With continuous shrinkage priors such as the GIGG the posteriors of lag groups remain non-zero with probability one (Hahn and Carvalho, 2015). This hampers the understanding of which indicators impact the cyclical component and thus the interpretability of results. Thresholding to exactly zero those lag groups that have little effect on y_t, i.e. essentially eliminating them from the model, makes it easier to interpret and communicate model outcomes. And it allows to draw inference about model uncertainty which can be pervasive in macroeconomic applications (Giannone et al., 2021; Huber et al., 2019; Cross et al., 2020; Kohns and Bhattacharjee, 2020). Recently, Mogliani and Simoni (2021) extend the adaptive group-lasso prior applied to Almon-lag MIDAS regressions to spike-and-slab variable selection for that purpose. Instead, we propose the use of a sparsification algorithm that is motivated by the perspective of a Bayesian decision maker who seeks the smallest subset of groups that best summarise the forecasts of the model

(1)-(3), inspired by Hahn and Carvalho (2015) and Woo and Owen (2019).[9] This can be achieved by minimising a utility function over the Euclidean distance between a linear model that penalises group-size akin to Zou (2006) and our model's prediction:

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

where $\tilde{\boldsymbol{Y}}$ refers to a realisation from the posterior predictive distribution $p(\tilde{\boldsymbol{Y}}|\boldsymbol{y}) = \int p(\tilde{\boldsymbol{Y}}|\boldsymbol{y}, \boldsymbol{Z^{(m)}}, \theta, \bullet)p(\theta|\boldsymbol{y}, \boldsymbol{Z^{(m)}}, \bullet)d\theta$, and $|| \bullet ||_p$ refers to the ℓ_p-norm.[10] Similar to the logic of adaptive group-lasso (Wang and Leng, 2008), the penalisation 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. The Bayes optimal solution for α is obtained by integrating out the posterior uncertainty from the predictive distribution, as well as in the parameters θ (Lindley, 1968) (see Appendix A.3).

In the following, we first show the analytical solution we derive for (8), and then discuss the assumptions needed to derive it. For a full derivation, see appendix A.3. The sparsified estimate $\alpha_k^{*(s)}$ for each Gibbs-sampling step $s = 1, \cdots, S$, is given by:

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

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 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 $T^{-1}\tilde{\boldsymbol{Z}}_{\boldsymbol{k}}^{\boldsymbol{(m)}\prime}\tilde{\boldsymbol{Z}}_{\boldsymbol{k}}^{\boldsymbol{(m)}} = \boldsymbol{I}$. This serves to simplify the solution, and as shown in (Simon and Tibshirani, 2012), not orthonormalising groups ignores the cross-correlation of group members in k such that the algorithm implicitly prefers to not threshold groups whose covariance is large, and also ignores that $\boldsymbol{Z}_k^{\boldsymbol{(m)}}$ might be on different scales.[11] Secondly, we make use of the work by Ray and Bhattacharya (2018); Chakraborty et al. (2020) who show that, when setting $\phi_k^{(s)} = \frac{1}{\theta_k^{(s)}}$, iterative solution methods such as the coordinate descent (Friedman et al., 2010), converge already after the first cycle. This gives us the analytical solution 9.

The relative frequency of high-frequency lag-group k selected in $\alpha^{*(s)}$ over all Gibbs draws

[9] Note that using the unsparsified posterior estimates of θ for prediction is already optimal in terms of empirical risk (Chakraborty et al., 2020), so that the main goal of sparsification is communication.

[10] Note that for simplicity we define the predictive distribution over in-sample values of $\boldsymbol{Z^{(m)}}$, but in principle any data can be used for the analysis.

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

will be used to report inclusion probabilities that inform on the relative impact of an indicator. See Woody et al. (2021) and Chakraborty et al. (2020) for formal justification of the model selection uncertainty and the asymptotic risk properties, respectively.

3 Data set and 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 stylised publication calendar. In the following, we outline the data set, the stylised 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 labour 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. 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 stabilised 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 prior to estimation.[12]

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.[13] 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.

[12] See Table B1 in the appendix for an overview of the data and their respective transformations.

[13] 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. Alternatively, we also employed the commonly used EM algorithm (Bańbura and Modugno, 2014) for interpolation, and we found that there is little difference in the sample under investigation.

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.[14]

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 stylised pseudo real-time data release calendar (see Table 1).

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, see also Table B1 in the appendix for a list of data series included. "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).

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 stylised 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.[15] For

[14] 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.

[15] 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).

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).[16] To draw samples from the predictive 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 real time root-mean-squared-forecast-error (RMSFE) which are calculated for each nowcast period as:

$$\text{RMSFE} = \sqrt{\frac{1}{T_{end}} \sum_{t=1}^{T_{end}} (y_{T+t} - \hat{y}^v_{T+t|\Omega^v_{T+t-1}})^2}, \tag{10}$$

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 real-time continuous rank probability score (CRPS):

$$\text{CRPS} = \frac{1}{T_{end}} \sum_{t=1}^{T_{end}} \frac{1}{2} \left| y_{T+t} - y^v_{T+t|\Omega^\nu_{T+t-1}} \right| - \frac{1}{2} \left| y^{v,A}_{T+t|\Omega^v_{T+t-1}} - y^{v,B}_{T+t|\Omega^v_{T+t-1}} \right|. \tag{11}$$

Note that in (11), $y^{v,A,B}_{T+j|\Omega^v_{T+j-1}}$ are 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 generalisation of the mean-absolute-forecast-error. To facilitate the discussion below, the objective in terms of predictive precision is to minimise both evaluation metrics.

4 Empirical Results

In section 4.1, we first develop intuition on the trend and cyclical components in UK GDP growth by focusing on posterior estimates of the various model components (1)-(3). Section

[16] This is in contrast to full system mixed frequency methods which represent missing low- and high-frequency information as unobervable state variables (Bańbura et al., 2013).

4.2, evaluates nowcast performance of the Trend-SV-t-BMIDAS model against alternative model specifications without the Trend and SV-t model features, or with alternative specifications for the BMIDAS component. Section 4.3 then unpacks the signals from the high-frequency indicators via inclusion probabilities derived via the sparsification algorithm.

4.1 Analysing UK GDP growth via the Trend-SV-t-BMIDAS

Figure 1 shows the posterior estimates of the cyclical (blue) and trend (orange) components (upper panels, separating the pre-pandemic period and the Covid-19 period) as well as the stochastic volatility components of GDP growth (lower left panel) and the trend equation (lower right panel) from the Trend-SV-t-BMIDAS model with GIGG prior. The cyclical component captures high frequency movements in GDP growth. It tracks the quarter-to-quarter movements in 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 GDP growth and can be interpreted as a time-varying long-run growth estimate. We observe a gradual slowdown in UK GDP growth since the early 2000s, with an additional temporary decrease in the trend during the Great Financial Crisis (GFC). Throughout the Covid-19 pandemic, the trend remains largely unchanged, hence the model interprets the extreme movements in GDP growth seen during the pandemic as transitory in nature.

Further, the t-distributed volatility estimate of the observation equation (lower left panel) shows a sharp and strong increase during the pandemic, by far exceeding the increase observed during the GFC. The model attributes the bulk in the increase in variance to be related to GDP growth itself and not to its long-run trend. However, trend volatility does gradually increase during the pandemic, even though credible intervals are wide, pointing to the possibility of a more persistent increase in the variance of long-run GDP growth. Although it remains an open question whether the Covid-19 pandemic has induced scarring in terms of repercussions to the long-run UK GDP growth trend, initial findings from our model suggest that the shock affected cyclical variation more than long-run trends.

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 Figure B2 in the appendix). But even prior to the pandemic, the inclusion of fat-tailed error variance is important to separate the slow-moving trend from cyclical movement. To elucidate this point, Figure 2 shows the posterior trend and cyclical component over the period until 2019Q4, from the baseline model that includes SV and t-distributed errors (blue lines and shaded areas), the model with SV but

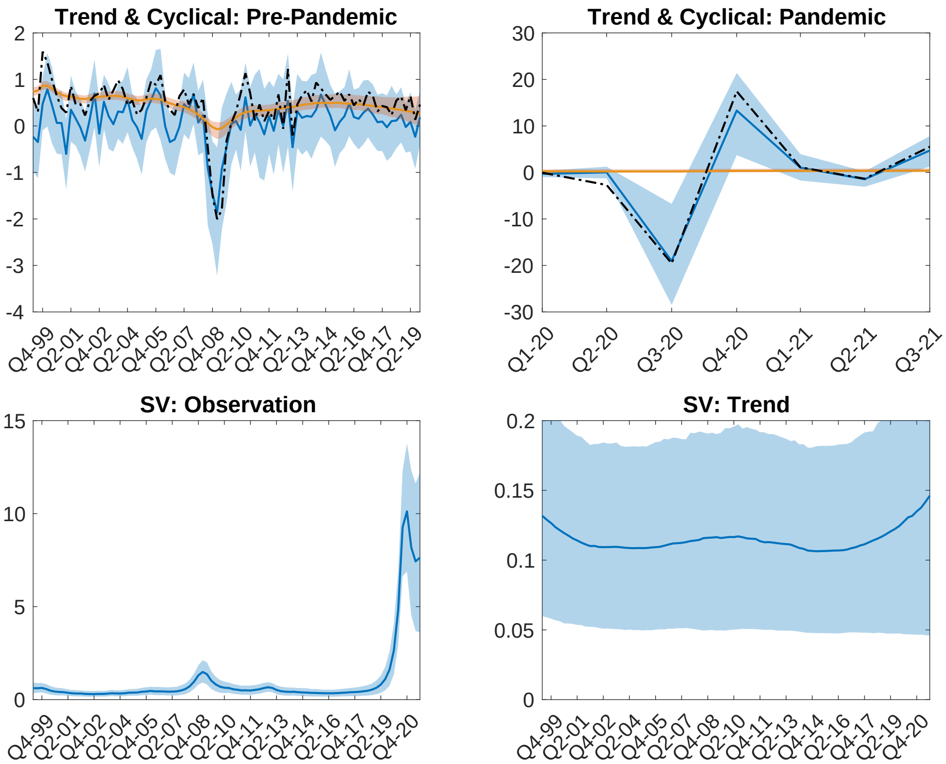


Figure 1: Posterior estimates for trend, cyclical component and stochastic volatilities.

Notes: Results from the T-SV-t-BMIDAS model with GIGG prior. The components are estimated over the full sample - pre-pandemic and pandemic cycle and trend are shown separately for readability. Orange lines and areas show the posterior means for the trends in GDP growth. Blue lines and areas show posterior medians of the cyclical components in GDP growth (upper panel) and stochastic volatilities in GDP growth (lower panel, left) and trend (lower panel, right). Shaded areas show 95% credible intervals.

no t-distr. errors (orange), as well as for constant variance only (green).[17] The model with SV shows the least intuitive results since it assigns most of the cyclical variation to the trend component, whereas the cycle remains very stable throughout the sample including during the GFC. Adding SV without t-distr. errors in a sample that contains both quieter periods and larger shocks such as the GFC likely results in overly discounting variation available to the cyclical component and instead over-fitting the trend. On the other hand, the baseline model achieves a sensible trend-cycle decomposition since it is able to separate large transitory shocks, that are handled by the fat tails, from smaller but more pervasive ones that are captured by the SVs in GDP growth and its trend. Both the cyclical and trend components are estimated with high precision in the baseline model. By contrast, in the constant variance model credible intervals are much larger, since the uncertainty of shocks permeates fully into the cyclical and trend components. While this model does assign short-term fluctuations to the cyclical

[17] The posterior estimates are based on the information set at the first nowcast period of 2019Q4. As section 4.2 will show, early nowcasts over the data release cycle are those benefiting the most from the inclusion of a time-varying trend with SV-t errors during the pre-pandemic period.

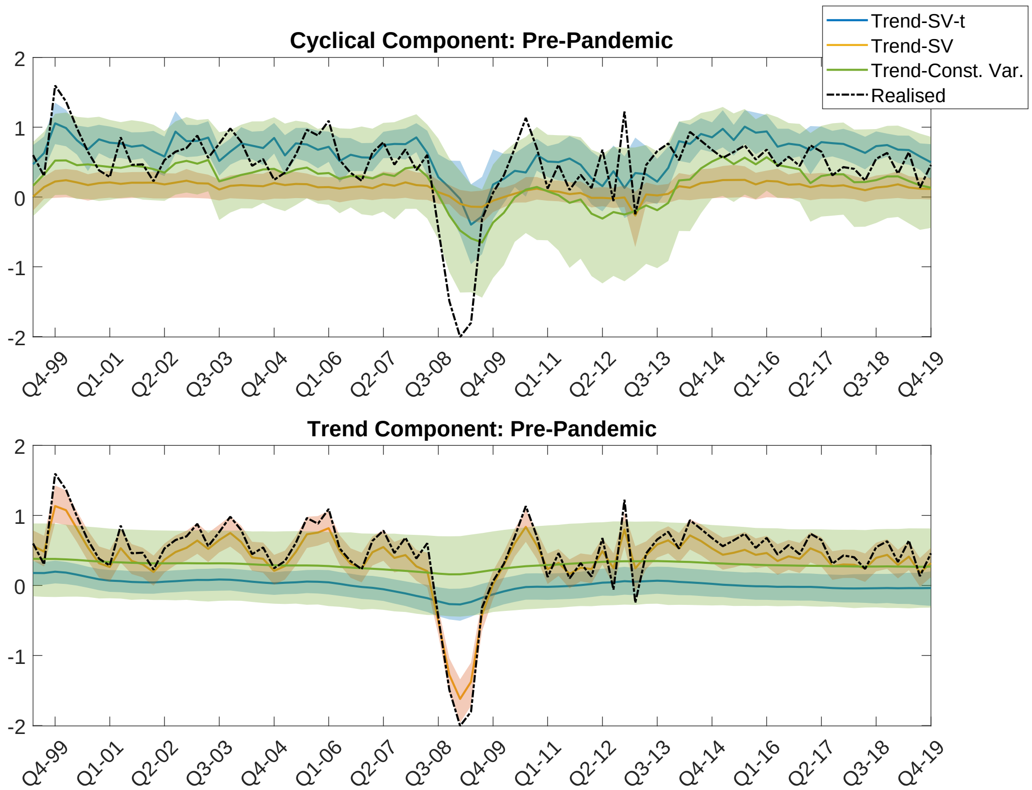


Figure 2: Posterior trend estimates from different BMIDAS specifications, pre-pandemic.
Notes: Posterior means of the trend up until 2019Q4 from the Trend-SV-t BMIDAS model (blue), the Trend-SV model (orange), and the Trend-Const.Var model (green). Estimation at the 1st nowcast period. Black dashed line shows realisation of real UK GDP growth. Shaded areas represent 95% credible intervals.

component, this goes at the expense of an almost unresponsive trend.

4.2 Nowcast evaluation and the role of model features

Having provided intuition for the posterior of the model components, we now evaluate the nowcast performance of the Trend-SV-t-BMIDAS model. First, we focus on the role of model features. Then, we evaluate the preferred Trend-SV-t model against a range of benchmarks, with a focus on performance during the pandemic.

4.2.1 Role of model features

Figure 3 assesses the role of model components by comparing the nowcast performance of the baseline model with versions where the new features, i.e. the time-varying trend, the stochastic volatilities, and their t-distributed errors, are shut off one by one or jointly. The figure shows root mean square forecast errors (RMSFE, upper panel) and continuous rank probability scores (CRPS, lower panel) over the data release cycle (days ahead of GDP release) on the x-axis.

Over the pre-pandemic period (left panel), the Trend-SV-t BMIDAS (blue solid line) clearly

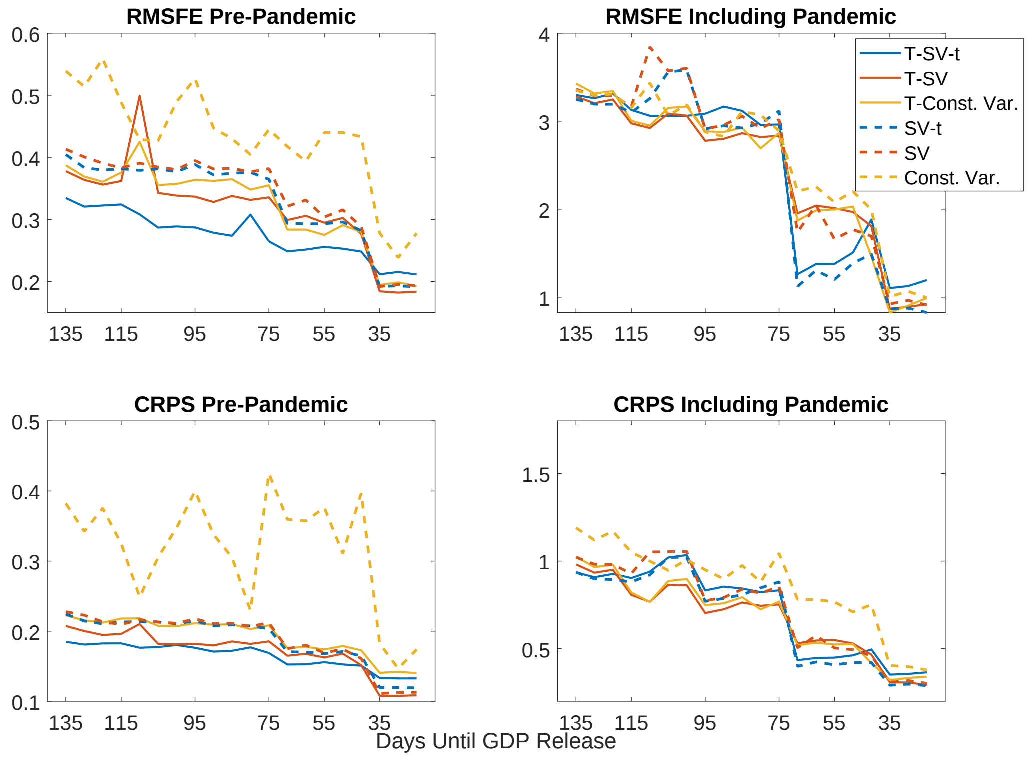


Figure 3: Nowcast performance for different BMIDAS model components.

Notes: Absolute root mean square forecast errors (RMSFE) and continuous rank probability scores (CRPS) over nowcast periods (days before GDP release on x-axis), for different specifications of BMIDAS with GIGG prior. Solid lines: models including a time-varying trend (T-). Dashed lines: models without trend. Blue lines: models with stochastic volatility with t-distr. errors (SV-t). Orange lines: models with stochastic volatility (SV). Yellow lines: models with constant variance. Results for pre-pandemic period evaluate nowcasts over 2011Q2-2019Q4, results including the pandemic evaluate nowcasts until 2021Q3.

outperforms the alternative model specifications in terms of point and density nowcasts for most of the nowcast periods (135-35 days prior to GDP release). The model's nowcast performance nearly continuously improves as new data come in, whereas some of the other models exhibit ups and downs in the nowcast performance. Sizeable improvements in nowcast performance are observed 105, 65 and 35 days prior to GDP release, which coincides with the releases of PMIs and of the production and service indicators for the first and second month of the reference quarter, respectively. Adding the time-varying trend component to the model (solid lines) provides benefits for point and density nowcasts prior the pandemic, independently of the specification of volatilities, in line with existing evidence for the United States by Antolin-Diaz et al. (2017). Also adding stochastic volatility combined with t-distr. errors improves the nowcast performance further and stabilises it over the data release cycle. As such, the nowcast performance of models with a trend but no t-distr. errors (orange and yellow solid lines) temporarily deteriorates when survey data for the first month of the reference quarter get released (110 days to GDP), whereas the performance of the baseline model improves with that release. The model with constant variance and no trend (yellow dashed lines) clearly performs worst, which underlines the importance of incorporating at least one of the proposed features into BMIDAS models—even over samples that exclude the Covid-19 pandemic.

When including the Covid-19 pandemic (right panel), not surprisingly, nowcast errors are much higher for all models, particularly early on in the data release cycle, where the episode of the initial downturn generated by lockdowns and the health crisis inflates nowcast errors. Differences across model variants are relatively small over most nowcast periods. Nowcasts gradually improve in the early parts of the data release cycle until the release of PMIs for the second month of the reference quarter (75 days prior to GDP), followed by a strong improvement in performance with the release of "hard" indicators pertaining to the first month (65 days prior to GDP). The nowcast performance of all models strongly improves at this point, but more so for models with stochastic volatility and t-distr. errors with and without trend (blue lines). Whether the time-varying trend is added or not makes less of a difference once the pandemic period is included, in line with the findings from Figure 2 that models that do not account for outliers struggle to identify the Covid-19 pandemic related downturn as temporary and instead over-fit the trend. Finally, 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. Such a model is not able to capture the large shift in the data neither via increased uncertainty nor through trend shifts.

Overall, we find that adding a time-varying trend to the BMIDAS model helps nowcast performance during the pre-pandemic period and does not detriment performance once the Covid-19 shock is included. We also find that adding stochastic volatilities is only useful when it is also combined with t-distributed errors—otherwise, the model over-fits the trend and shows a weaker and more volatile performance. Based on these results, for the rest of the paper, we choose the Trend-SV-t BMIDAS model as our preferred specification.

4.2.2 Evaluation against alternatives to the BMIDAS component

Next, we assess the nowcasting performance of the proposed Trend-SV-t-BMIDAS model with GIGG prior against the following benchmark models.

- **AR(2)**: represents a purely auto-regressive benchmark, does not include trends, stochastic volatility or t-errors.
- **Combination**: Univariate MIDAS regressions for each of the K high frequency macroeconomic indicators, estimated with a normal prior and combined according to their discounted RMSFE and CRPS performance akin to Stock and Watson (2004). We follow Carriero et al. (2019) by setting the discount factor $\delta = 0.95$. For comparability with our model, we estimate the univariate MIDAS regressions with Trend-SV-t components.

- **MF-DFM**: Similar to Antolin-Diaz et al. (2021), this dynamic factor model also includes Trend-SV-t components, but it captures the cyclical information via a dynamic factor that exploits co-movement across the K indicators contamporaneously and at up to two lags. The priors on the latent state components are similar to the ones outlined in section 2.2.3.
- **Trend-SV-t-BMIDAS-HS**: model 1-3 with the horseshoe prior, thus another flexible shrinkage prior, but without group-shrinkage. This represents a benchmark in terms of the prior for the Trend-SV-t BMIDAS.[18]

Figure 4 plots RMSFE and CRPS over the data release cycle, and Table 2 shows the evaluation metrics of the proposed model and benchmarks relative to the AR model, on average across all nowcast periods and for selected periods that correspond to releases of PMIs and real activity indicators. Stars indicate significantly different point and density nowcasts as per Diebold et al. (1998). Three main findings emerge. First, similarly to many other studies (e.g. Foroni et al. (2015); Carriero et al. (2015, 2019)), models that exploit monthly macroeconomic indicators outperform the autoregressive model, with relative RMSFE and CRPS against the AR(2) lying clearly and significantly below 1 for most models throughout the nowcast cycle. The relative advantage against the AR(2) model becomes even more pronounced when including the pandemic quarters into the evaluation, albeit the differences become insignificant for individual nowcast quarters due to the large uncertainty during that shock.

Second, the proposed Trend-SV-t model with GIGG prior is among the strongest models prior to the pandemic and outperforms all other models once pandemic quarters are included in the evaluation. Prior to the pandemic, the MIDAS Combination (dashed lines) and DFM (dashed-dotted lines) show the strongest performance in terms of density nowcasts, but the Trend-SVt-BMIDAS with GIGG prior performs similarly to these in terms of point forecasts. The Combination displays the least variation over nowcast periods given that it averages out nowcast errors, but it is outperformed by the DFM and Trend-SVt-BMIDAS with GIGG for later nowcast periods. Whereas the nowcast performance of the other models stagnates during later nowcast cycles when moving closer to GDP release, the proposed model shows continuous improvements, indicating that it can exploit real activity data releases particularly well. Indeed, once real activity indicators for the first month are published 35 days prior to GDP release, the relative RMSE (CRPS) of the proposed model significantly improves against the AR(2) by 50% (40%). The relative advantage of the proposed Trend-SVt-BMIDAS is even stronger when including Covid-19 observations into the evaluation sample: it considerably outperforms the

[18] We also compared nowcasting performance to BMIDAS models with other frontier prior choices such as the adaptive group-lasso with spike and slab prior model for Almon lag transformed data as proposed by Mogliani and Simoni (2021). We find that our models are again competitive or outperform these benchmarks. These results are available on request.

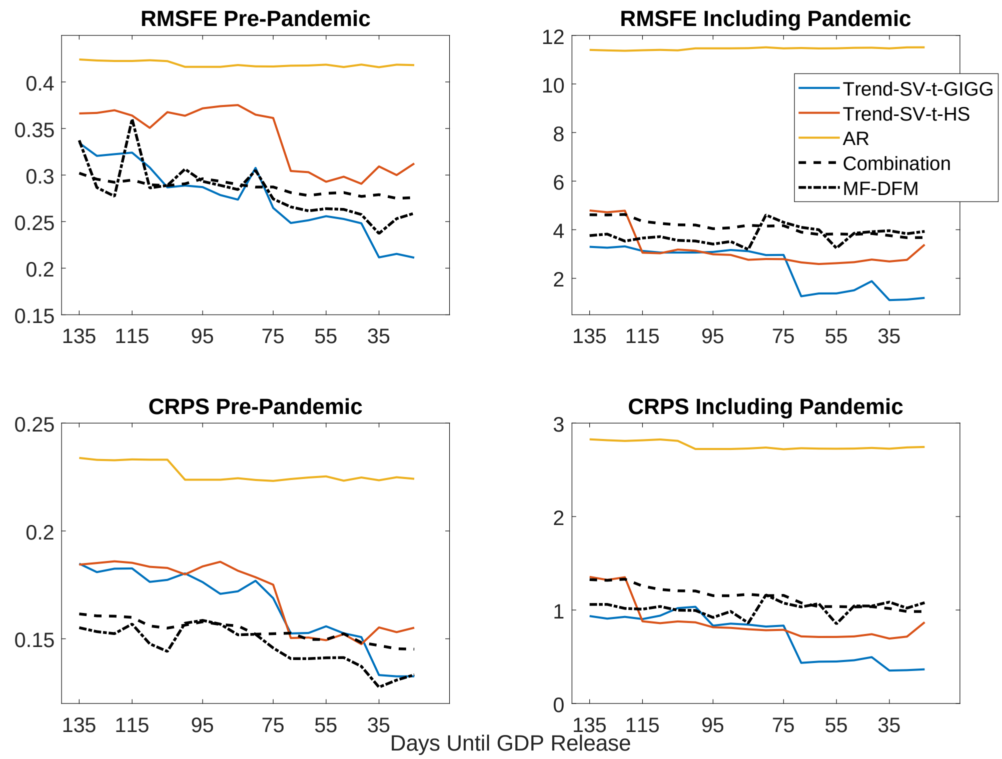


Figure 4: Nowcast performance of Trend-SV-t-BMIDAS with GIGG compared to benchmarks.
Notes: Absolute RMSFE and CRPS over nowcast periods (days before GDP release on x-axis) for Trend-SV-t-BMIDAS with GIGG prior (blue lines), Trend-SV-t-BMIDAS with HS prior (orange), Combined U-BMIDAS (black dashed), MF-DFM (black dotted-dashed), and AR(2) (yellow). Results for pre-pandemic period evaluate nowcasts over 2011Q2-2019Q4, results including the pandemic evaluate nowcasts until 2021Q3.

Table 2: Relative Evaluation Metrics

Nowcast Periods	Average	6	13	18	Average	6	13	18
	RMSFE Pre-Pandemic				RMSFE Including Pandemic			
AR	**0.42**	**0.42**	**0.42**	**0.42**	**11.45**	**11.4**	**11.47**	**11.48**
MF-DFM	0.67***	0.75	0.67**	0.63**	0.32***	0.33	0.32	0.34
Combination	0.68***	0.68**	0.67**	0.67**	0.36***	0.37	0.34	0.33
Trend-SV-t-HS	0.81***	0.87	0.73	0.74	0.28***	0.28	0.23	0.24
Trend-SV-t-GIGG	0.66***	0.68**	0.60*	0.51**	0.21***	0.27	0.11	0.10
	CRPS Pre-Pandemic				CRPS Including Pandemic			
AR	**0.23**	**0.23**	**0.22**	**0.22**	**2.76**	**2.83**	**2.72**	**2.73**
MF-DFM	0.63***	0.60**	0.64**	0.60**	0.36***	0.36	0.36	0.36
Combination	0.68***	0.66**	0.68*	0.66**	0.41***	0.43	0.39	0.37
Trend-SV-t-HS	0.75***	0.79	0.67*	0.70	0.32***	0.31	0.26	0.26
Trend-SV-t-GIGG	0.73***	0.76	0.68*	0.60*	0.26***	0.36	0.16	0.13

Notes: The table shows the average RMSFE and CRPS 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). RSMFE and CRPS 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)

Combination and DFM approach throughout the data release cycle, but again the relative gain is strongest for later nowcast periods.

Third, regarding the role of the shrinkage prior, exerting group-level shrinkage and taking into account the high-frequency correlation structure via the GIGG prior (blue line) has preferable performance compared to the horseshoe (HS) prior (yellow line). Prior to the pandemic, the model with GIGG prior has lower point nowcast errors throughout and almost always better density fit. When including the pandemic quarters, it is strongest for the first nowcast periods and again starting from 75 days ahead of GDP publication. As we illustrate via variable inclusion probabilities in section 4.3, this nowcast gain is achieved because the GIGG prior shrinks the information set towards a more sparse selection of indicators, and because it shifts its signal extractions towards indicators related to the service sector when the pandemic hits.

4.2.3 Nowcasts during the Covid-19 pandemic

At which point during the pandemic and the data release cycle does the proposed Trend-SVt BMIDAS model with GIGG prior achieve its forecast gains? Figure 5 visualises the nowcasts over time for the proposed model compared to the sets of models discussed in section 4.2.2, for selected nowcast periods. Pre-pandemic, nowcasts are fairly close to each other. The Trend-SV-t BMIDAS nowcasts show somewhat more volatility, which can explain their slightly weaker performance compared to Combined MIDAS and DFM. For the Covid-19 pandemic quarters, unsurprisingly, all models miss the large unprecedented trough early in the data release cycle. However, the Trend-SV-t BMIDAS with GIGG prior is the quickest to update nowcasts to the trough and subsequent rebound in activity, and its nowcasts in later nowcast periods are closest to the actual realisation. In early nowcast periods, it is the only model to indicate the large rebound in activity for Q3-2020. And once the real activity indicators for the first reference month have been published, it shows the largest downward adjustment for Q2-2020. The other models are less responsive, or belatedly nowcast a trough without capturing the recovery.[19]

For our proposed model, 95% credible intervals illustrate the role of the SV-t feature for the uncertainty around nowcasts during the Covid-19 pandemic. The initially wide credible intervals show the expected increase in nowcast uncertainty around the trough and recovery of GDP growth, but also that uncertainty decreases substantially after Q3-2020 for later nowcast periods. Hence, with more "hard" macroeconomic information, the model indicates a return toward reduced uncertainty after the Covid-19 shock.

[19] A likely explanation for the weaker performance of the MIDAS Combination model is that the combination weights are slow to adapt in the standard discounted weights approach, and allocate the increased GDP growth variation to the outlier component. Similarly, the DFM may struggle due to fixed loadings which in normal times load heavily on production surveys, but these were less informative during the pandemic.

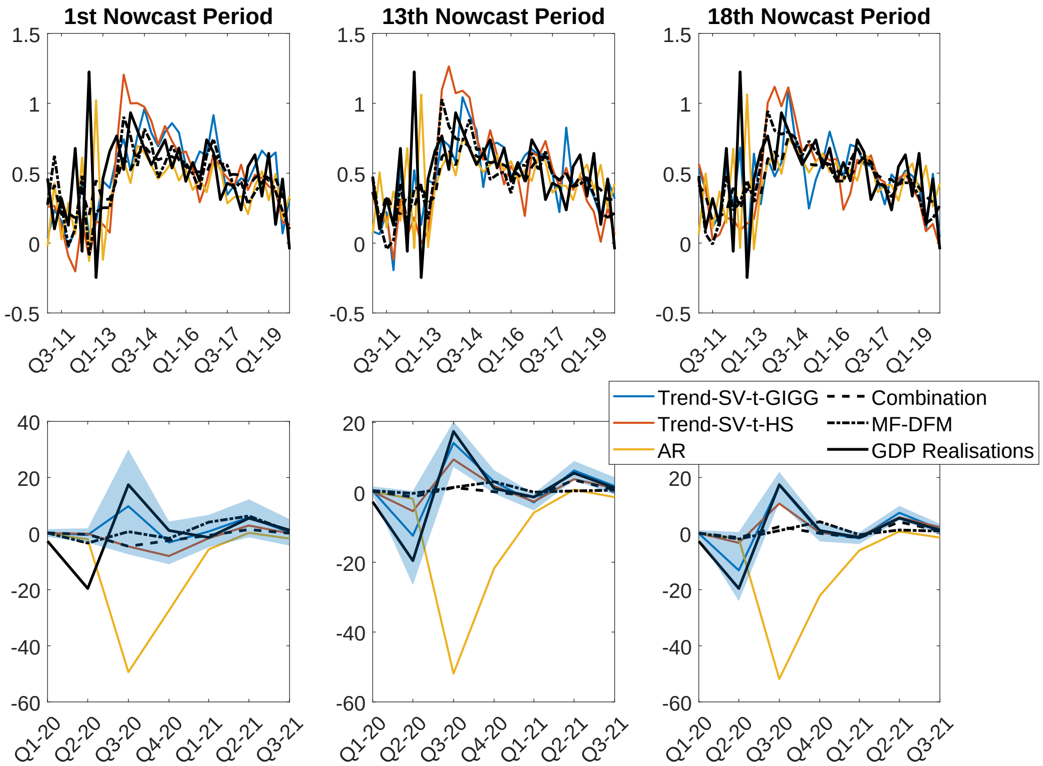


Figure 5: Posterior mean and density nowcasts, selected nowcast periods.

Notes: Upper panel shows nowcasts over time for the period 2011Q1 to 2019Q4 (x-axis) at different points in the data release cycle (columns), lower panel shows nowcasts over the pandemic quarters on the x-axis. Models are the same as in Figure 4. Shaded areas refer to Trend-SV-t-GIGG model and show 95% credible interval. Black solid lines show quarterly GDP growth realisations.

4.3 Interpreting Signals via Variable Inclusion Probabilities

The group-variable selection achieved via the sparsification algorithm (9) equips us with an intuitive way to communicate signals within a Bayesian, decision theoretically based setting. Variable selection is communicated as the inclusion probability of high-frequency lag group k, i.e. a macroeconomic indicator and its lags, into the linear model (8) over posterior samples. The higher the inclusion probability of lag group k, the larger its impact onto the predictions of the model. Inclusion probabilities turn out rather stable over the sample period before the Covid-19 pandemic, so that we focus on averages over sub-samples.

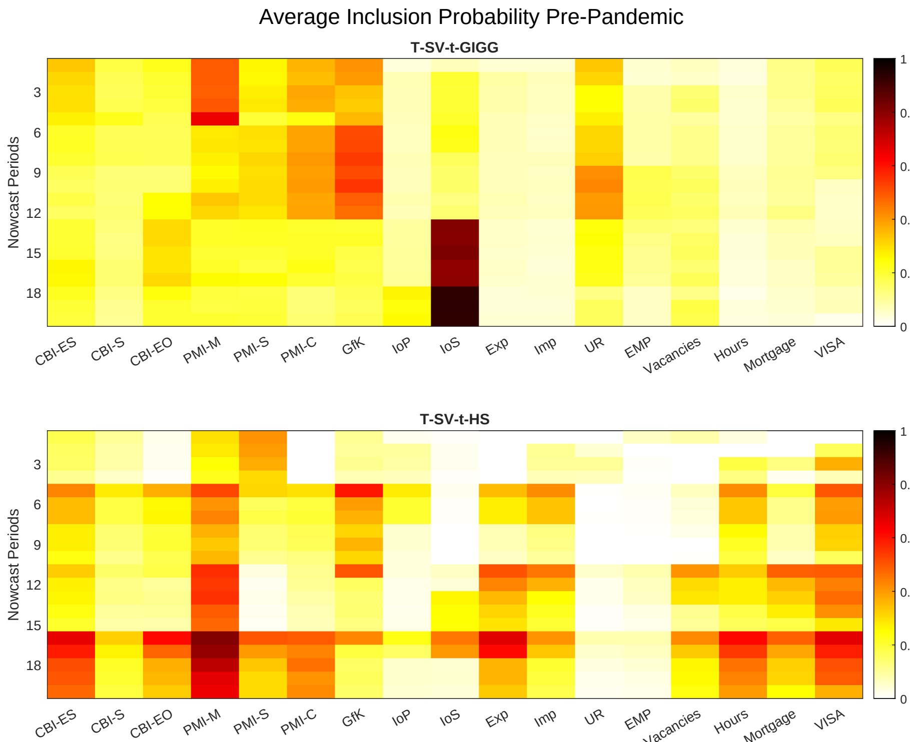
Figure 6 presents heatmaps for inclusion probabilities for each indicator (x-axis) over nowcast periods (y-axis). The lower sub-plots in each panel shows corresponding results using the horseshoe prior without group-shrinkage. Pre-pandemic (panel a), the model with GIGG prior (upper sub-plot) selects a sparse specification, as indicated by only a few indicators shaded in red. The model uses signals from one to three indicators at a time, shifting to signals from other indicators as they get released. Early in the data release cycle, the model mostly exploits a few survey variables: Manufacturing and Construction PMIs for very early nowcasts and then GfK consumer confidence for the first month of the reference quarter, once it gets released in nowcast period 5. Labour market data also plays a role when released in period 9. However, once the Once, however, "hard" real economic information get published 65 days prior to GDP release (period 13), the model loads almost exclusively on the index of services. This agrees with earlier findings that survey indicators provide the main early signals for quarterly GDP, but that incoming hard information becomes more important for nowcasts later in the data release cycle (Bańbura et al., 2013; Carriero et al., 2015; Anesti et al., 2017).

When the pandemic is included into the sample (panel b), both prior specifications make the model read a wider range of signals as indicated by the overall darker colours. Nonetheless, the model with GIGG prior remains much more sparse, with a clear pattern of exploiting different signals over the data release cycle. It keeps exploiting survey indicators in the early part of the sample, but interestingly the most informative indicator now becomes the Service PMI, followed by Construction PMI. Over nowcast periods 5 to 12, signal from mortgage lending are relevant too. Little focus is put on the GFK and labour market data. Once hard economic data become available, very strong signals are read from the index of services, and additional signals ahead of GDP release come from the Service PMI and index of production. In this, the model efficiently exploits that during the pandemic most of the disruption to the economy was stemming from lockdowns affecting the service sector as well as initially a shut-down of the housing and construction sectors, whereas consumer confidence and manufacturing remained much less affected and labour market data were distorted by the furlough scheme. This likely helps the model with GIGG prior to capture the recovery from the Covid-19 induced trough early in the data release cycle and to update its nowcast of the initial trough in GDP earlier compared to other models, as discussed in section 4.2.3.

On the other hand, the model with horseshoe prior (lower sub-plots) shows dense inclusion pattern both pre-pandemic and particularly once the pandemic is include. It draws on signals from surveys, real, labour and personal finance indicators in a diffused way over the data release cycle. The denser signal extraction with the horseshoe prior could stem from increased cross-sectional correlation in the face of large macroeconomic shocks, documented in studies such as Rockova and McAlinn (2021); McAlinn et al. (2018); Frühwirth-Schnatter and Lopes (2018). However, given that the Covid-19 shock affected specific sectors more than others, a dense solution can represent a disadvantage. And the heavy reliance of the model with horseshoe prior on surveys even when hard economic information is available can explain the relatively weak performance of the model discussed in section 4.2.2.

Overall, this illustrates how group shrinkage helps exploit sparse signals from highly correlated macroeconomic data. In this, our findings provide new impetus to the debate on the "illusion of sparsity", where models with a dense cyclical component have been found to fore-

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



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

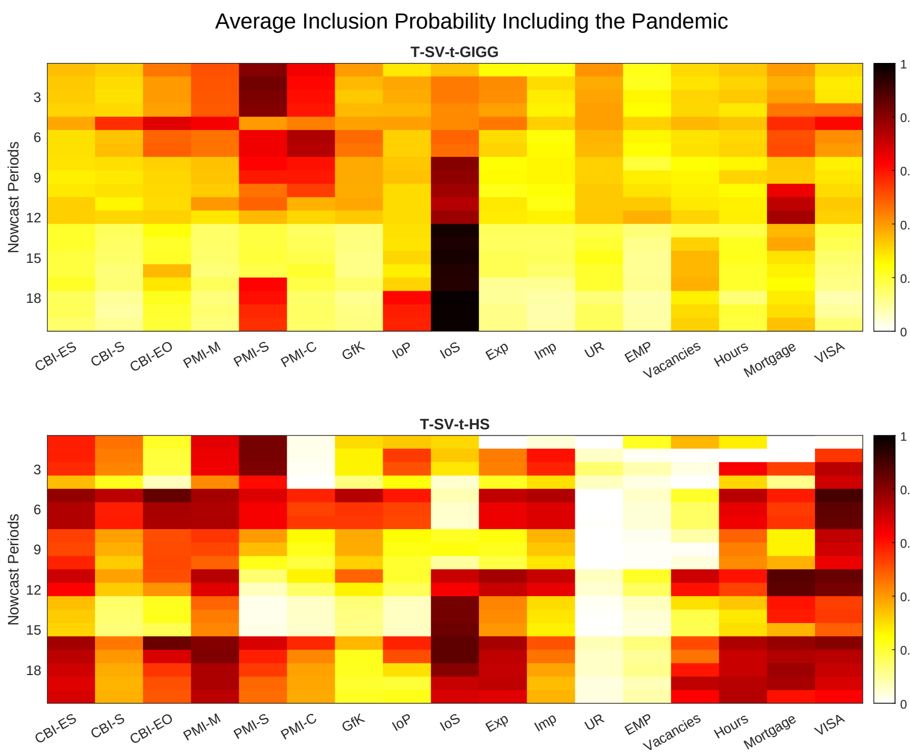


Figure 6: Average posterior inclusion probabilities over high frequency lags for each indicator.
Notes: Heatmaps show nowcast periods on the y-axis, darker colour indicates higher cumulative posterior inclusion probabilities of all high frequency lags of an indicator. Upper panel a) shows results for evaluation sample prior to the pandemic, lower panel b) including the pandemic. Sub-plots show different prior specifications of the Trend-SV-t BMIDAS model: GIGG with Almon lag restrictions (GIGG-AL), GIGG with U-MIDAS lags (GIGG-UM), and horseshoe prior with Almon lag restrictions (HS-AL).

cast better in applications with macroeconomic data than models that prefer sparse model solutions (Giannone et al., 2021; Fava and Lopes, 2021; Kohns and Bhattacharjee, 2020; Cross et al., 2021). We find, however, that denser specifications, as the horseshoe prior model, or the MF-DFM, do not necessarily forecast better than the group-sparse model with GIGG prior, particularly in presence of the Covid-19 shock affecting the economy heterogeneously. Similar results for forecasting applications have been found for comparable aggressive shrinkage priors to the horseshoe in Fava and Lopes (2021) who show elevated variable selection uncertainty with strongly correlated data (Piironen et al., 2020).

4.4 Sensitivity Analyses

We conduct two sensitivity analyses, one with respect to the stability of nowcast performance of our preferred model when holding coefficient fixed over the Covid-19 pandemic, and the other using an unrestricted MIDAS specification instead of Almon lag polynomials. Evaluation results in comparison to the baseline specification are presented in Figure (7).

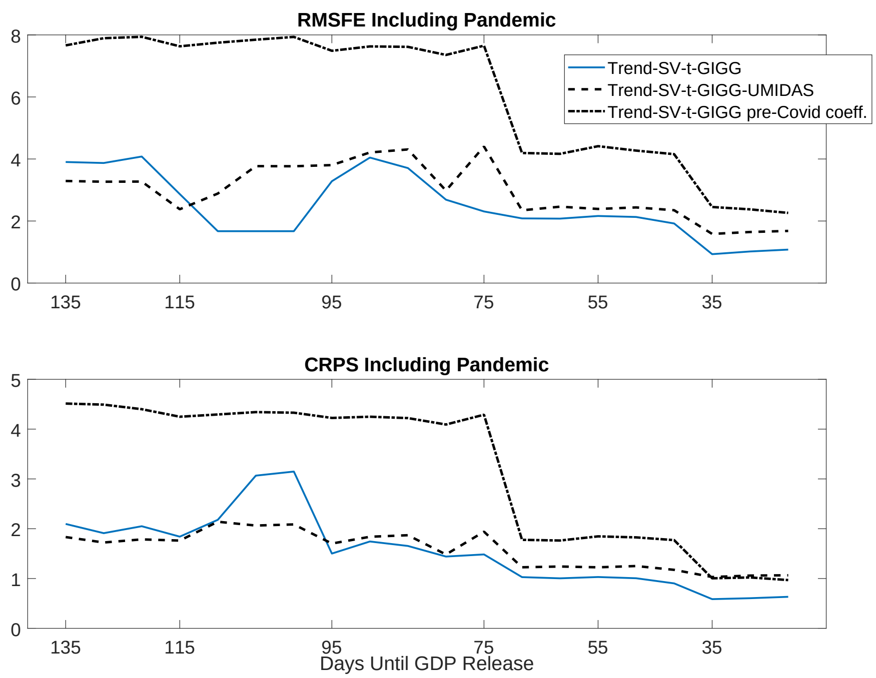


Figure 7: Nowcast performance of alternative specifications of the Trend-SV-t-BMIDAS.
Notes: The Trend-SV-t-GIGG-UMIDAS uses U-MIDAS sampled data (Foroni et al., 2015) instead of Almon restricted lags. The Trend-SV-t-GIGG pre-Cov.coeff nowcasts Q1-2020 onwards with coefficients based on the sample that ends with Q4-2019, and nowcasts evaluated over four quarters, Q3-2020 to Q2-2021.

Holding model coefficients fixed over the Covid-19 pandemic is motivated by Schorfheide and Song (2021) who show that MF-VAR models without flexible error components can achieve similar nowcast performance (after the trough) by simply omitting the problematic first two quarters of 2020 in the estimation sample, which is similar to using scale processes that inversely move with periods of large volatility, thereby downweighting their effect. We nowcast with our model starting Q1-2020 with parameter coefficients based on the sample ending with Q4-2019. If the GDP dynamics during the Covid-19 shock truly were only outliers in the UK, the nowcast performance for Q3-2020 to Q2-2021 should be similar to those with the original Trend-SV-t-GIGG model. However, we find that using pre-pandemic coefficients, the model performs significantly worse both in point as well as density nowcasts up until about 40 days until GDP release, and then shows similar performance. As discussed in section 4.3, with onset of the pandemic, the baseline model with GIGG prior shifts variable selection towards indicators reflecting sectors that were hit strongly by the shock. These dynamics are not picked up when the Covid-19 period is omitted from the estimation.

In a second sensitivity exercise, we relax the assumption of a restricted MIDAS structure via Almon-polynomial distributed lags. The Trend-SV-t-BMIDAS with unrestricted MIDAS (U-MIDAS) structure has comparable performance, but does somewhat worse during later nowcast periods. This indicates that the additional regularisation via Almon-polynomials helps nowcast performance when the dimension of the information set increases, but also that the GIGG prior adapts well enough to the U-MIDAS data to produce comparable performance to the baseline model. Results for inclusion probabilities using a U-MIDAS structure are comparable to the baseline (available upon request), indicating that the sparse specification stems from imposing the GIGG group-shrinkage prior and not from Almon-polynomial restrictions.

5 Conclusion

In this paper, we have proposed a new Bayesian MIDAS framework, the T-SV-t-BMIDAS model combined with a flexible group-shrinkage prior and a sparsification step for variable selection motivated by Bayesian decision theory. In an application of the model to nowcasting UK GDP growth, we have shown that 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 it largely draws signals from "hard" indicators rather then survey data.

Two important insights regarding the role of model features and prior choice emerge. First, adding a long-run trend or t-distributed stochastic volatility into a multivariate BMIDAS model substantially improves predictive performance. Second, the GIGG shrinkage prior enhances performance by inducing group-wise sparsity while enabling the model to flexibly shift between signals. In our application, this feature proves particularly relevant once the Covid-19 pandemic is included into the analysis, since flexibly shifting across signals from sparse groups of indicators, rather than relying on dense signals due to aggregate macroeconomic co-movement, can be advantageous in presence of a large heterogeneous shock.

The proposed framework is flexible, as it combines model features and nests various models as special cases. The account for time-varying long-run trends allows to detect trend shifts quickly. Results are interpretable since variable inclusion probabilities can be communicated over time and over forecast horizons. All this makes an attractive tool not only for nowcasting GDP growth, but also for instance for medium-term GDP or inflation forecasting. The group-shrinkage prior has potential to efficiently regularise information towards a sparse set of signals from a range of other data sets, such as disaggregated price or labour market data, where we can expect groupings both among lags of observations as well as across sectors. 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.

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.

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.

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.

Bhadra, A., J. Datta, N. G. Polson, B. Willard, et al. (2019). Lasso meets horseshoe: A survey. *Statistical Science 34*(3), 405–427.

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.

Carriero, A., A. B. Galvão, and G. Kapetanios (2019). A comprehensive evaluation of macroeconomic forecasting methods. *International Journal of Forecasting 35*(4), 1226–1239.

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. (2017a). Notes on Bayesian macroeconometrics. Manuscript available at http://joshuachan. org.

Chan, J. C. (2017b). The stochastic volatility in mean model with time-varying parameters: An application to inflation modeling. *Journal of Business & Economic Statistics 35*(1), 17–28.

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.

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

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.

Cross, J. L., C. Hou, G. Koop, and A. Poon (2021). Macroeconomic Forecasting with Large Stochastic Volatility in Mean VARs. BI Norwegian Business School.

Cross, J. L., C. Hou, and A. Poon (2020). Macroeconomic forecasting with large Bayesian VARs: Global-local priors and the illusion of sparsity. *International Journal of Forecasting*.

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

Devroye, L. (2014). Random variate generation for the generalized inverse gaussian distribution. *Statistics and Computing 24*(2), 239–246.

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.

Durbin, J. and S. J. Koopman (2002). A simple and efficient simulation smoother for state space time series analysis. *Biometrika 89*(3), 603–616.

Fava, B. and H. F. Lopes (2021). The illusion of the illusion of sparsity: An exercise in prior sensitivity. *Brazilian Journal of Probability and Statistics 35*(4), 699–720.

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. (1994). Data augmentation and dynamic linear models. *Journal of Time Series Analysis 15*(2), 183–202.

Frühwirth-Schnatter, S. and H. F. Lopes (2018). Sparse Bayesian factor analysis when the number of factors is unknown. *arXiv preprint arXiv:1804.04231*.

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.

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. and M. Marcellino (2018). *Applied economic forecasting using time series methods*. Oxford University Press.

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.

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.

Hörmann, W. and J. Leydold (2014). Generating generalized inverse Gaussian random variates. *Statistics and Computing 24*(4), 547–557.

Huber, F., G. Koop, and L. Onorante (2019). Inducing sparsity and shrinkage in time-varying parameter models. *arXiv preprint arXiv:1905.10787*.

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*.

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

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 (2020). Developments on the Bayesian Structural Time Series model: trending growth. *arXiv preprint arXiv:2011.00938*.

Koop, G. and L. Onorante (2019). Macroeconomic Nowcasting Using Google Probabilities. *Topics in Identification, Limited Dependent Variables, Partial Observability, Experimentation, and Flexible Modeling: Part A (Advances in Econometrics 40*, 17–40.

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.

McAlinn, K., V. Rockova, and E. Saha (2018). Dynamic sparse factor analysis. *arXiv preprint arXiv:1812.04187*.

McCausland, W. J., S. Miller, and D. Pelletier (2011). Simulation smoothing for state–

space models: A computational efficiency analysis. *Computational Statistics & Data Analysis 55*(1), 199–212.

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.

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., 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.

Polson, N. G., J. G. Scott, and J. Windle (2014). The Bayesian bridge. *Journal of the Royal Statistical Society: Series B: Statistical Methodology*, 713–733.

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

Rockova, V. and K. McAlinn (2021). Dynamic variable selection with spike-and-slab process priors. *Bayesian Analysis 16*(1), 233–269.

Schorfheide, F. and D. Song (2021). Real-time forecasting with a (standard) mixed-frequency VAR during a pandemic. NBER Working Paper No 29535.

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 (2004). Combination forecasts of output growth in a seven-country data set. *Journal of Forecasting 23*(6), 405–430.

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.

Van Der Pas, S. L., B. J. Kleijn, A. W. Van Der Vaart, et al. (2014). The horseshoe estimator: Posterior concentration around nearly black vectors. *Electronic Journal of Statistics 8*(2), 2585–2618.

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

Woloszko, N. (2020). Tracking activity in real time with Google Trends. OECD Working Paper No 1634.

Woo, J. and A. L. Owen (2019). Forecasting private consumption with Google Trends data. *Journal of Forecasting 38*(2), 81–91.

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.

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

A Additional Model Descriptions

A.1 Posteriors

A.1.1 Further Insight into the GIGG Posterior

In order to highlight the effect of of different choices on a_k and b_k, we re-write the posterior in (6) in terms of its shrinkage coefficient representation:

$$\begin{aligned}\overline{\theta}_{k,j} &= \Lambda_*((X'\Lambda_{t,h}^{-1}\boldsymbol{z}_{\boldsymbol{k}}^{(\boldsymbol{m})})^{-1} + \Lambda)\hat{\theta}_{k,j} \\ &= (1-\kappa_{k,j})\hat{\theta}_{k,j},\end{aligned} \tag{12}$$

where we define $\boldsymbol{z}_{\boldsymbol{k}} = (z'_{k,1}, \cdots, z'_{k,T})'$, $\Lambda_{t,h} = diag(\lambda_1^2 e^{h_1}, \cdots, \lambda_T^2 e^{h_T})$ and $\hat{\theta}_{k,j} = (\boldsymbol{z}_{\boldsymbol{k}}^{(\boldsymbol{m})\prime}\Lambda_{t,h}^{-1}\boldsymbol{z}_{\boldsymbol{k}}^{(\boldsymbol{m})})^{-1})\boldsymbol{z}_{\boldsymbol{k}}^{(\boldsymbol{m})\prime}\Lambda_{t,h}^{-1}\tilde{\boldsymbol{y}}$ can be viewed as a conditional maximum likelihood estimate for $\theta_{k,j}$. We suppress the indication of monthly frequency (m) for convenience here. Note that for this representation, we have assumed that the maximum likelihood estimate for group k exists and that the $p_k + 1$ lags within $\boldsymbol{z}_{\boldsymbol{k}}$ have been orthogonalised. Since the group-size is likely to be much smaller than the sample size, and given that we will group-orthogonalise the the data anyway (see section 2.3), these only represent very mild assumptions. Under these assumptions, it is easy to verify that the shrinkage coefficient $\kappa_{k,j|\bullet} = \frac{1}{1+\tilde{s}_{k,j}^2\vartheta^2\gamma_k^2\varphi_{k,j}^2}$ is bounded between 0 and 1 and thus dictates how far away the prior shrinks the coefficients from the maximum likelihood solution. It is easy to see that $\vartheta^2\gamma_k^2\varphi_{k,j}^2 \to \infty$, $\overline{\theta}_{k,j} \to \hat{\theta}_{k,j}$. The distribution $\pi(\kappa_{k,j})$ which is implicitly defined via the priors on γ_k^2 and $\varphi_{k,j}^2$ determine the a-priori shrinkage behaviour we can expect. By assuming $\gamma_k^2\varphi_{k,j}^2 \sim \beta'(a_k, b_k)$, the joint distribution for $\boldsymbol{\kappa}_{\boldsymbol{k}}$ can be factored as:

$$\begin{aligned}\pi(\kappa_k|\bullet) =& \frac{\Gamma(a_k + (p_k+1)b_k)}{\Gamma(a_k)\Gamma(b_k)^{k+1}} \prod_{j=1}^{p_k} \tilde{s}_{k,j}^{b_k} \left(1 + \sum_{j=1}^{p_k} \tilde{s}_{k,j} \frac{\kappa_{k,j}}{(1-\kappa_{k,j})}\right)^{-(a_k+(1+p_k)b_k)} \times \\ & \left(\prod_{j=1}^{p_k} \kappa_{k,j}^{b_k-1}(1-\kappa_{k,j})^{-(b_k+1)}\right),\end{aligned} \tag{13}$$

where $\tilde{s}_{k,j} = s_{k,j}\sum_{t=1}^{T}\frac{1}{\lambda_t}e^{-h_t}$ and $s_{k,j}$ is the jth lag's variance. This joint distribution factors into a dependent part, influenced by a_k, and an independent part, determined by b_g. Plot (B1) further elucidates this behaviour, which shoes the joint-shrinkage distribution for a group-size of 2.

As expected, when b_g is relatively small compared to a_g, then the sparsity level enforced by a_g dominates: lower left hand panel showcases a situation in which relatively little shrinkage is exerted because a_g is relatively large compared to b_g, while the upper right hand panel's joint distribution is characterised by a independently distributed, very extreme horseshoe behaviour

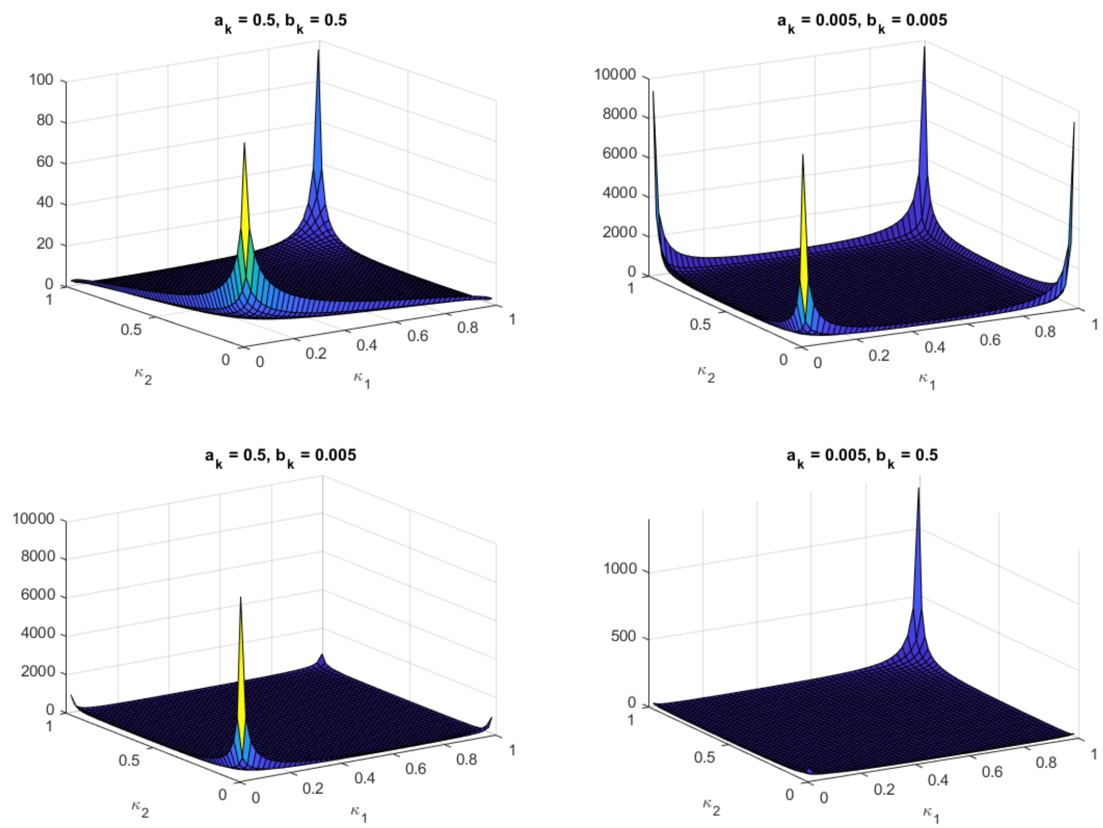


Figure B1: Bi-variate shrinkage coefficient plots for various hyper-parameter values. a_g controls group-level sparsitym while b_g controls the degree of correlation with the overall sparsity level.

(the U shape is much narrower than that implied by the standard horseshoe). When b_g instead is relatively large compared to a_g, then the shrinkage behaviour will tend to be symmetric. The exact group-horseshoe case ($a_g = b_g$), for example, resembles a joint U-shape. Hence, for the exact group-horseshoe, a-priori either the entire group will be shrunk to zero, or all coefficients are left relatively un-perturbed. This exposition will help understand which behaviour to expect given the choice of the hyper-parameters for the nowcast application in the empirical application.

A.1.2 Posteriors of Hyper-parameters

The deviations of the conditional posteriors for $\vartheta, \gamma_k^2, \varphi_{kj}$ for $k = 1, \cdots, K$ and $j = 1, \cdots, p_k + 1$ immediately follow from the presentation in Boss et al. (2021). Following Boss et al. (2021), we employ a mixture representation of the β' prior via an inverse-gamma distributed auxiliary variable ν_p. The conditional posteriors $\vartheta, \gamma_k^2, \varphi_{kj}^2, \nu_p$ are thus proportional to:

$$(\vartheta|\boldsymbol{y}, \bullet) \sim IG(\frac{\sum_{k=1}^{K}(p_k + 1) + 1}{2}, \theta' \Lambda_p^{-1} \theta / 2 + \frac{1}{\nu_p}) \tag{14}$$

$$(\gamma_k^{-2}|\boldsymbol{y},\bullet) \sim GIG(\frac{p_k+1}{2} - a_k, \frac{1}{\vartheta^2}\sum_{j=1}^{p_k+1}\frac{\theta_{kj}^2}{\varphi_{kj}^2}) \tag{15}$$

$$(\varphi_{kj}|\boldsymbol{y},\bullet) \sim IG(b_k + \frac{1}{2}, 1 + \frac{\theta_{kj}^2}{2\vartheta^2\gamma_k^2}) \tag{16}$$

$$(\nu_p|\boldsymbol{y},\bullet) \sim IG(1, \frac{1}{\vartheta}), \tag{17}$$

where GIG refers to the generalised inverse Gaussian distribution (Hörmann and Leydold, 2014) which we generate from using the efficient algorithm of (Devroye, 2014).

A.1.3 Posteriors of the State Space

In this section, we will detail the conditional posteriors of each of the remaining parameters of model which will be used to construct the Gibbs sampler in. For convenience, we reproduce the main Trend-SV-t model of section 2.1 here again:

$$\begin{aligned} &y_t = \tau_t + \theta' Z_{t-h}^{(m)} + \sqrt{\lambda_t} e^{\frac{1}{2}(h_0 + w_h \tilde{h}_t)} \tilde{\epsilon}_t^y, \\ &\tilde{\epsilon}_t^y \sim N(0,1),\ \lambda_t \sim IG(\nu/2, \nu/2) \end{aligned} \tag{18}$$

$$\begin{aligned} &\tau_t = \tau_{t-1} + e^{\frac{1}{2}(g_0 + w_g \tilde{g}_t)},\ \tilde{\epsilon}_t^g \sim N(0,1) \\ &\tilde{h}_t = \tilde{h}_{t-1} + \tilde{\epsilon}_t^h,\ \tilde{\epsilon}_t^h \sim N(0,1),\ \tilde{h}_0 = 0 \\ &\tilde{g}_t = \tilde{g}_{t-1} + \tilde{\epsilon}_t^g,\ \tilde{\epsilon}_t^g \sim N(0,1),\ \tilde{g}_0 = 0 \end{aligned} \tag{19}$$

$$\begin{aligned} &w_g \sim N(0, V_{w_g}),\quad w_h \sim N(0, V_{w_h}) \\ &h_0 \sim N(a_{0,h}, b_{0,h}),\quad g_0 \sim N(a_{0,g}, b_{0,g}) \\ &\tau_0 \sim N(a_{0,\tau}, b_{0,\tau}) \end{aligned} \tag{20}$$

As mentioned in the main body of the text, the state space components $(\boldsymbol{h}, \boldsymbol{g})$ are written in their non-centred form. The non-centred form of a state space allows to dissect the latent processes into a time-varying part $(w_h\tilde{\boldsymbol{h}}, w_g\tilde{\boldsymbol{g}})$ and a constant part (h_0, g_0), and this apply different amounts of shrinkage to each part. In doing so, one is also able to model the state standard deviations (w_h, w_g) as part of the conditional mean of the state equations. This allows to exert more shrinkage than with tranditional variance priors such as the inverse-gamma Frühwirth-Schnatter and Wagner (2010); Chan (2017a). The centred state space can be recovered by replacing:

$$
\begin{aligned}
h_t &= h_0 + w_h \tilde{h}_t \\
g_t &= g_0 + w_g \tilde{g}_t
\end{aligned} \tag{21}
$$

Posterior of τ

To derive the joint posterior of $\boldsymbol{\tau}$, we make use of the methods proposed by Chan and Jeliazkov (2009); McCausland et al. (2011) as they enable sampling of all states $(\tau_1, \cdots, \tau_T)$ simultaneously. Compared to more traditional forward-sampling-backwards-smoothing algorithms of Carter and Kohn (1994); Durbin and Koopman (2002) which sample states one time-step at a time, this represents an improvement in statistical efficiency as well as computational efficiency. The computational efficiency comes from the special band-matrix form of the resultant state posterior which allows for the use of very efficient sparse matrix operations (Chan and Jeliazkov, 2009).

Since we only need the conditional posterior $\pi(\boldsymbol{\tau}|\tau_0, w_\tau, \bullet)$, we proceed similar to the exposition in 6, by defining the relevant conditional likelihood based on $\boldsymbol{y^*} = \boldsymbol{y} - \theta \boldsymbol{Z}^{(m)}$, hence the observations of the target, accounted for the cyclical component. It follows from 18 that then:

$$
\boldsymbol{y}|\boldsymbol{\tau}, \bullet \sim N(\boldsymbol{\tau}, \Lambda_h \Lambda_t^{1/2}). \tag{22}
$$

To derive the implicit prior on $\boldsymbol{\tau}$, start by vectorising the state process $\boldsymbol{\tau}$ in 19:

$$
H\tau = \tilde{\alpha}^{\boldsymbol{\tau}} + \eta^\tau, \tag{23}
$$

where $\boldsymbol{\tau} = (\tau_1, \cdots, \tau_T)'$, $\tilde{\alpha}^\tau = (\tau_0, 0, \cdots, 0)'$, $\eta^\tau \sim N(0, \Lambda_g)$, $\Lambda_g = diag(e^{g_1}, \cdots, e^{g_T})$ and H is the first difference matrix. From 23, one can write the joint prior of $\boldsymbol{\tau}$ as:

$$
\boldsymbol{\tau}|\tau_0, \bullet \sim N(\tau 1_T, (H'\Lambda_g^{-1} H)^{-1}), \tag{24}
$$

where 1_T is a column of ones with dimension $T \times 1$. Since the all priors in the model are a-priori independent, the conditional posterior is found by standard calculations:

$$
\boldsymbol{\tau}|\boldsymbol{y}, \tau_0, \bullet \sim N(\hat{\boldsymbol{\tau}}, K_\tau^{-1}), \tag{25}
$$

where $K_\tau = H'\Lambda_g^{-1} H + \Lambda_h^{-1} \Lambda_t^{-1/2}$, $\hat{\boldsymbol{\tau}} = K_\tau^{-1}(H\Lambda_g^{-1} H \tau_0 1_T + \Lambda_h^{-1} \Lambda_t^{-1/2} y^*)$. Due to the special band matrix structure on the posterior, we can significantly speed up computation time by using sparse matrix computations.

To sample τ_0, recall that it only appears in $\tau_1 = \tau_0 + e^{\frac{1}{2}(g_0 + w_g * \tilde{g}_1)} \tilde{\epsilon}_1^\tau$. Hence, assuming with the independent prior in 20, the posterior is:

$$
\tau_0|\boldsymbol{y}, \bullet \sim N(\hat{\tau}_0, K_{\tau_0}^{-1}), \tag{26}
$$

$K_{\tau_0} = \frac{1}{b_{0,\tau}} + \frac{1}{e^{g_1}}$, $\hat{\tau}_0 = K_{\tau_0}^{-1}(\frac{a_{0,\tau}}{b_{0,\tau}} + \frac{\tau_1}{e^{g_1}})$.

Posterior of $\boldsymbol{h}$

To derive the posterior of $\boldsymbol{h}$, we use the commonly employed approximate discrete mixture sampler of Kim et al. (1998). Define $y_t^+ = log((y_t - \tau_t - \theta' Z_{t-h}^{(m)})/\sqrt{\lambda_t})^2$, $\tilde{\epsilon}_t^y = log(\tilde{\epsilon}_t^y)^2$, then the relevant conditional likelihood for h reduces to:

$$\boldsymbol{y}^+ = h_0 1_T + w_h \tilde{\boldsymbol{h}} + \tilde{\epsilon}^{y^+}. \tag{27}$$

Since the error distribution $\tilde{\epsilon}^{y^+}$ now lives on the log-scale, the standard normal regression results cannot be directly applied to this conditional likelihood. The distribution now follows a $log\chi_1^2$ distribution. Instead, we follow Kim et al. (1998) by introducing component indicators $s = (s_1, \cdots, s_T)$ such that given these, $(\tilde{\epsilon}^{y^+}|s) \sim N(d_s, \Omega_s)$ where d_s and Ω_s are obtained from a 7-point Gaussian mixture approximation to the $log\chi_1^2$. See Kim et al. (1998) for definitions of d_s and Ω_s. Conditional on d_s and Ω_s, the likelihood and prior for $\tilde{h}$ become normal again, so that the logic for deriving the posterior of τ from above can be re-applied. The prior for $\tilde{h}$ becomes $N(0, (H'H)^{-1})$, so that the posterior is rendered:

$$\tilde{\boldsymbol{h}}|\boldsymbol{y}, \bullet \sim N(\hat{\tilde{\boldsymbol{h}}}, K_{\tilde{h}}^{-1}), \tag{28}$$

$K_{\tilde{h}} = H'H + w_h^2 \Omega_s^{-1}$, $\hat{\tilde{\boldsymbol{h}}} = K_{\tilde{h}}^{-1}(w_h \Omega_s^{-1}(\boldsymbol{y}^+ - h_0 1_T - d_s))$.

The remaining conditional posteriors associated with $\boldsymbol{h}$ are those of h_0 and w_h. Notice from 27, conditional on $\tilde{\boldsymbol{h}}$, d_s and Ω_s, the joint posterior of h_0 and w_h can be found using simple regression results. Define $X_h = (1_T', \tilde{\boldsymbol{h}}')'$, $\zeta_h = (h_0, w_h)$, $a_{0,\zeta_h} = (a_{0,h}, 0)$, $b_{0,\zeta_h} = diag(b_{0,h}, V_{w_h})$, then $\zeta_h \sim N(a_{0,\zeta_h}, b_{0,\zeta_h})$. The joint posterior is:

$$\zeta_h|\boldsymbol{y}, \bullet \sim N(\hat{\zeta}_h, K_{\zeta,h}^{-1}), \tag{29}$$

$K_{\zeta_h} = (b_{0,\zeta_h}^{-1} + X_h' \Omega_s^{-1} X_h)$ and $\zeta_h = K_{\zeta_h}^{-1}(b_{0,\zeta_h}^{-1} a_{0,\zeta_h} + X_h' \Omega_s^{-1}(\boldsymbol{y}^+ - d_s))$. We set the rather uninformative hyper-priors $a_{0,h} = 0, b_{0,h} = 10, V_{w_h} = 0.1$.

Posterior of $\boldsymbol{g}$

The posterior of $\boldsymbol{g}$ is similarly derived to that of $\boldsymbol{h}$ by replacing: $\boldsymbol{y}^+ = log(H\boldsymbol{\tau})^2$, $\tilde{\epsilon}^{y^+} = log(\tilde{\epsilon}^\tau)$, $a_{0,h} = a_{0,g}, b_{0,h} = b_{0,g}, V_{w_h} = V_{w_g}$. We set $a_{0,g} = 0$, $b_{0,g} = 10, V_{w_g} = 0.1$.

Posterior of λ and ν

To derive the posterior $p(\boldsymbol{\lambda}|\bullet)$, notice that each λ_t is univariate and independently distributed. Hence:

$$p(\boldsymbol{\lambda}|\bullet) \propto \prod_{t=1}^{T} \lambda_t^{-\frac{\nu+1}{2}+1} e^{-\frac{1}{2\lambda_t}(\nu + \frac{(y_t - \tau_t - \theta' Z_{t-h}^{(m)})^2}{exp(h_t)})}. \tag{30}$$

Notice that these are kernels of the inverse-Gamma distribution:

$$\lambda_t \sim \mathcal{G}^{-1}(\frac{\nu+1}{2}, \frac{1}{2}\frac{(y_t - \tau_t - \theta' Z_{t-h}^{(m)})^2}{exp(h_t)}) \tag{31}$$

Finally, regarding the unknown ν, we specify a uniform prior $\nu \sim U[2, 50]$. The lower limit ensures that the variance σ_y^2 exists, and 50 is chosen to be reasonably large such that the upper limit generates an error variance close to a normal. The conditional posterior boils down to:

$$\begin{aligned} p(\nu|\bullet) &\propto p(\boldsymbol{\lambda} p(\nu)) \\ &\propto \prod_{t=1}^{T} \frac{(\nu/2)^{\frac{\nu}{2}}}{\Gamma(\nu/2)} \lambda_t^{-(\frac{\nu}{2}+1)} e^{-\frac{\nu}{2\lambda_t}} \\ &= \frac{(\nu/2)^{\frac{T}{2}}}{\Gamma(\nu/2)^T} (\prod_{t=1}^{T})^{-(\frac{\nu}{2}+1)} e^{\frac{\nu}{2}\sum_{t=1}^{T}\lambda_t^{-1}} \end{aligned} \tag{32}$$

where the first definition follow from the fact that the priors are independent. This distribution is non-standard. To sample from this distribution, we make use of an independent Metrolpolis-Hastings within Gibbs sampling step.

By slight abuse of notation, define the target density as f, the current state of the Markov chain as X and the proposal state as Y, then the proposal Y is accepted with probability

$$\alpha(X, Y) = min\left\{\frac{f(Y)g(X)}{f(X)g(Y)}, 1\right\}, \tag{33}$$

where $g(.)$ is the proposal density. In order for the Metropolis-Hastings sampler to quickly explore the typical set of $\nu|\bullet$, g should be close to f. To ensure this, we define g as a normal with mean equal to the mode of f and covariance equal to the negative Hessian evaluated at the mode. To find the mode, we use the Newton-Raphson method. The Hessian is analytically available (Chan, 2017a).

A.2 Sampling Algorithm

In order to estimate the Trend-SV-t-BMIDAS model with GIGG prior, we make use of a Metropolis-within-Gibbs sampler. With the posterior distributions described in sections A.1.2,A.1,6, in hand, we sequentially sample from the following posterior distributions:

1. Sample $\theta|\bullet \sim p(\theta|\boldsymbol{y}, \bullet)$
2. Sample hyper-parameters $\vartheta, \gamma_k^2, \varphi_{kj}^2, \nu_p$ in one block
   (a) $\vartheta^2 \sim p(\vartheta^2|\boldsymbol{y}, \bullet)$
   (b) $\gamma_k^2 \sim 1/p(\gamma_k^{-2}|\boldsymbol{y}, \bullet)$
   (c) $\varphi_{kj}^2 \sim p(\varphi_{kj}^2|\boldsymbol{y}, \bullet)$

(d) $\nu_p \sim p(\nu_p|\boldsymbol{y}, \bullet)$

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 $\sim p(w_g|\boldsymbol{y}, \bullet)$
6. Sample $\{\lambda_t\}_{t=1}^T \sim p(\lambda_t|\boldsymbol{y}, \bullet)$
7. Sample $\nu_p \sim p(\nu_p|\boldsymbol{y}, \bullet)$ with a Metropolis step as described after equation 32

We iterate sampling steps 1.-7. initially for 5000 times for burn-in and retain further 5000 samples for our analysis. To speed up the computations, we make use of the state sampling techniques of Chan and Jeliazkov (2009) and Bhattacharya et al. (2016). The former allows drawing steps 3.-5. in a non-recursive fashion which increases efficiency and can be sped up substantially using sparse-matrix operations. The latter speeds up computation and aids mixing when $\sum_{k=1}^{K}(p_k+1) >> T$, which is the case when using U-MIDAS samples data for the empirical application. For a discussion of this algorithm, see Bhattacharya et al. (2016).

A.3 Group-Selection Algorithm

This section gives further details on the derivation of group-variable selection algorithm in 8. For convenience, we replicate the objective function here again, omitting notation involving m for clarity:

$$\mathcal{L}(\tilde{\boldsymbol{Y}}, \alpha) = \frac{1}{2}||\boldsymbol{Z}\alpha - \tilde{\boldsymbol{Y}}||_2^2 + \sum_{k=1}^{K} \phi_k ||\alpha_k||_2, \tag{34}$$

For simplicity, assume that the predictions $\tilde{\boldsymbol{Y}}$ can be decomposed as $\tilde{\boldsymbol{Y}} = \boldsymbol{Z}\theta + \tilde{\boldsymbol{\epsilon}}^{\boldsymbol{y}}$, $\tilde{\boldsymbol{\epsilon}} \sim N(0, \Sigma)$. Intuitively, objective function 8 pushes those α_k to zero which have little influence on the predictions of our model, $\tilde{\boldsymbol{Y}}$.

We take the expectation with respect to 1) the expected risk and 2) $\theta|\boldsymbol{y}$ to account for all sources of uncertainty of the model[B1]:

$$\begin{aligned}\mathcal{L}(\tilde{\boldsymbol{Y}}, \alpha) &= E_{\tilde{\boldsymbol{Y}}|\bullet}[\tilde{\boldsymbol{Y}}, \alpha|\bullet] \\ &= \sum_{k=1}^{K} \phi_k ||\alpha_k||_2 + \frac{1}{2}||\boldsymbol{Z}\alpha - \boldsymbol{Z}\theta||_2^2 + \frac{1}{2}tr(\Sigma)\end{aligned} \tag{35}$$

Then, taking the expectation with respect to $\theta|\boldsymbol{y}$:

[B1] Since α is independent of Σ, the integration of posterior uncertainty of Σ results in a constant and thus does not further influence the optimisation problem

$$\mathcal{L}(\theta,\alpha) = E_{\theta|\boldsymbol{y}}[\mathcal{L}(\tilde{\boldsymbol{Y}},\alpha)]$$
$$= \frac{1}{2}||\boldsymbol{Z}\alpha - \boldsymbol{Z}\overline{\theta}||_2^2 + \frac{1}{2}tr(\boldsymbol{Z}'\boldsymbol{Z}\Sigma_\theta) + \sum_{k=1}^{K}\phi_k||\alpha_k||_2, \tag{36}$$

where $\Sigma_\theta = cov(\theta)$ and $\overline{\theta} = E(\theta)$. Dropping all constant terms, the objective function reduces to:

$$\mathcal{L}(\theta,\alpha) = \frac{1}{2}||\boldsymbol{Z}\alpha - \boldsymbol{Z}\overline{\theta}|| + \sum_{k=1}^{K}\phi_k||\alpha_k||_2. \tag{37}$$

Notice, that we follow Chakraborty et al. (2020); Huber et al. (2019) by solving 36 on a Gibbs iteration bases (instead over the average of the posterior). Traditional solution methods such as the coordinate descent (Friedman et al., 2010) iteratively solve the sub-gradients L times for each group k until convergence:

$$\alpha_k^l = (||r_k^{(l-1)}||_2^2 - \phi_k)_+ \frac{r_k^{(l-1)}}{||r_k^{(l-1)}||_2^2},\ l = 1,\cdots,L, \tag{38}$$

where r_k is the partial residual based on the previous iteration, $r^{(l)} = \boldsymbol{Z}_k'(\boldsymbol{y} - \boldsymbol{Z}_{-k}\alpha_{-k}^{(l-1)})$ and $-k$ refers to all but the k^{th} group. Orthonormalising $\boldsymbol{Z}_k$ via its SVD decomposition (which can conveniently be adapted as in (Breheny and Huang, 2015) when $p_k + 1 >> T$), and stopping the coordinate descent after one iteration as per Ray and Bhattacharya (2018); Chakraborty et al. (2020) such that $||r_k|| = ||\theta_k||_2^2$ results immediately in 9.

A.4 Almon-Lags Restricted MIDAS

This section gives further details on the restrictioned entailed by the restrictions proposed in Almon (1965) and Smith and Giles (1976).

Define $X_{k,t}$, for $k \in \{1,\cdots,K\}$ as the $(L+1\times 1)$ vector of high frequency lags which span the months of the reference quarter and any further lags. In the empirical application we consider months 0 to 5, such that $L+1=6$.

Assume for simplicity that $y_t = \sum_{k=1}^{K}\beta_k' X_{k,t} + \epsilon_t,\ \epsilon_t N(0,\sigma^2)$[B2]. Now, any linear restrictions on the β_k processes can equivalently be represented as linear transformations on the $X_{k,t}$, via a $(p_k + 1\times L)$ weighting matrix, Ξ_k. The notation in the main text can therefore be recovered as: $z_{k,t} = \Xi_k X_{k,t}$. When the weights are left unrestricted, as in U-MIDAS (Foroni et al., 2015) estimation, $\Xi_k = I_{L_k}$. In this case $\theta_k = \beta_k$.

Almon lag polynomial restrictions, proposed for economic prediction models by Almon (1965), confine the regression problem to $\beta_{k,l} = \theta_{k,i}l^i$ for lags $l = (1,\cdots,L_k)$ and param-

[B2] Notice that we suppress the notation indicating that the covariate set is in monthly frequency for readability eters of the polynomial $i = (0, \cdots, p_k)$. For a third degree polynomial and $L_k + 1 = 6$, $\beta_k = (\theta_0, \sum_{i=0}^{p_k} \theta_i, \sum_{i=0}^{p_k} \theta_i 2^i, \sum_{i=0}^{p_k} \theta_i 3^i, \sum_{i=0}^{p_k} \theta_i 4^i, \sum_{i=0}^{p_k} \theta_i 5^i)'$. With this, the free parameters are reduced from L_k to p_k for each k, which when $p_k << L_k$ induces parsimony. This, however, comes a the cost of bias, that may be non-vanishing with increasing sample size Andreou et al. (2010). These restrictions imply the $(i+1)^{\text{th}}$ row of Ξ_k equal to $[0^i, \cdots, L_k^i]$.

For many time-series forecasting applications, it makes intuitive economic sense to assume that the effect of far lags on the target are 0, and that the weights peter out to 0 on a smooth fashion across lags. Dubbed "end-point" restrictions, these are again linear restrictions and, thus, lead to straightforward manipulations of Ξ_k, see Smith and Giles (1976). Say, there ω restrictions (in the empirical application $\omega = 2$) then this reduces the free parameters to $1 + p_k - \omega$ for each k.

B Additional results

B.1 Further Detail on Data

Table B1: Further Details on Macro Data

Name	Acronym	Frequency	Transformation	Source
CBI: Volume of exp. Trades	CBI-ES	m	0	FAME
CBI: Volume of rep. Sales	CBI-S	m	0	FAME
CBI: Volume of exp. Output	CBI-EO	m	0	FAME
PMI: Manufacturing	PMI-M	m	0	FAME
PMI: Services	PMI-S	m	0	FAME
PMI: Construction	PMI-C	m	0	FAME
GfK Cons. Confidence	GfK	m	0	FAME
Index of Production	IoP	m	3	FAME
Index of Services	IoS	m	3	FAME
Exports	Exp	m	3	FAME
Imports	Imp	m	3	FAME
Unemployment Rate	UR	m	0	FAME
Employment	Emp	m	3	FAME
Job Vacancies	Vacancies	m	3	FAME
Hours Worked	Hours	m	3	FAME
Mortgage Approvals	Mortgage	m	3	FAME
VISA consumer spending	VISA	m	3	FAME
Real quarterly GDP growth (qoq)	GDP	q	3	FAME

Notes: The table shows the data used for the empirical application along with respective sampling frequencies (m = monthly, q = quarterly), tranformation applied (0 = no transformation, 1 = logs , 2 = first difference, 3 = growth rates) and data source. All data are downloaded from the UK data provider FAME. Please see the FAME website for further details on the data.

B.2 In-Sample Description

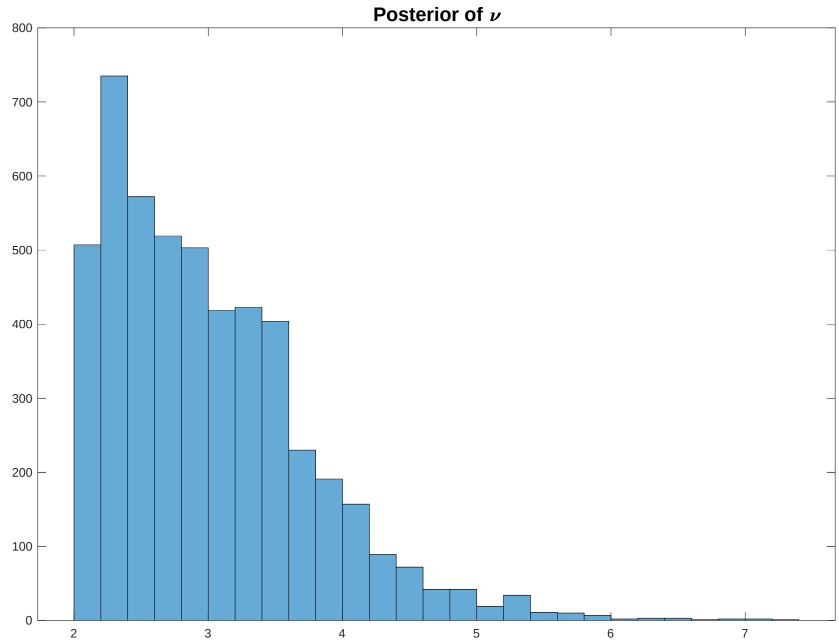


Figure B2: Posterior Degrees of Freedom of the t-distribution, based on the entire in-sample and last nowcast period.

Figure B2 has large mass over small posterior degrees of freedom for the t-distribution assumed for the errors of the observation equation of model [reference of equation 1]. The smaller the degrees of freedom, the more leptokurtic the tails of the observation equation's error distribution. Figure B2 shows strong identification of fat-tails.

B.3 Nowcast Performance GIGG Models

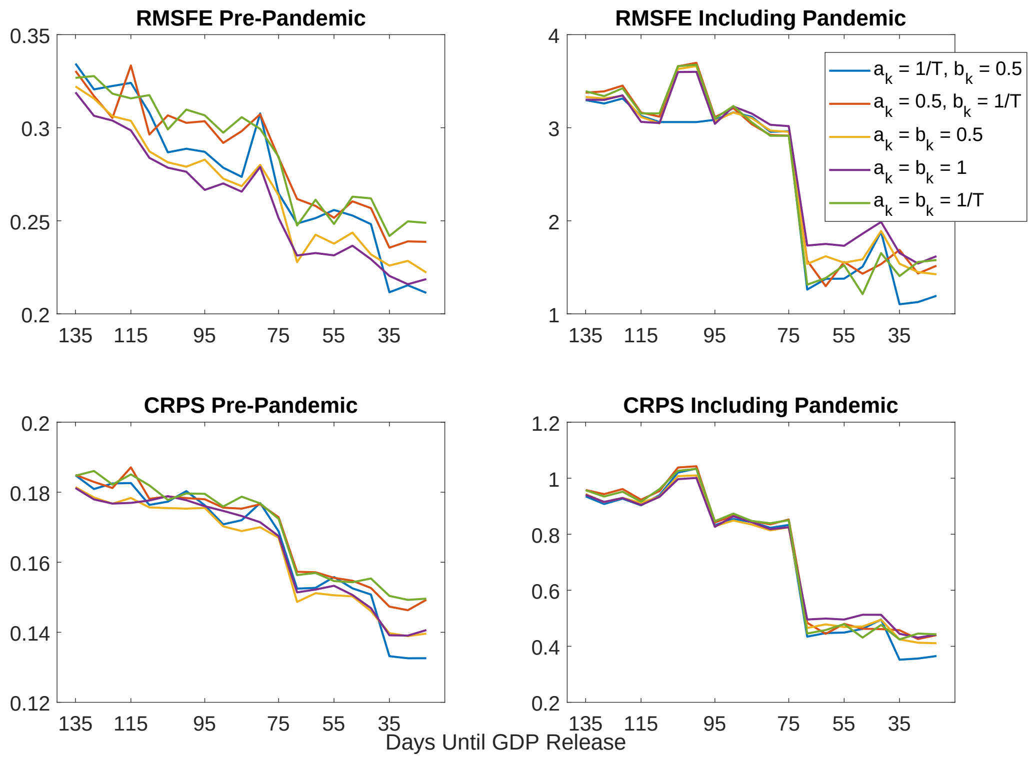


Figure B3: RSMFE and CRPS pre-pandemic and including pandemic for the Trend-SV-t-GIGG model with different hyper-parameter combinations. $1/T$ is adjusted to the in-sample length at each quarter the nowcasts are conducted. The GIGG prior implemented in the empirical application has $a_k = 1/T, b_k = 0.5$.

B.4 Inclusion Probabilities

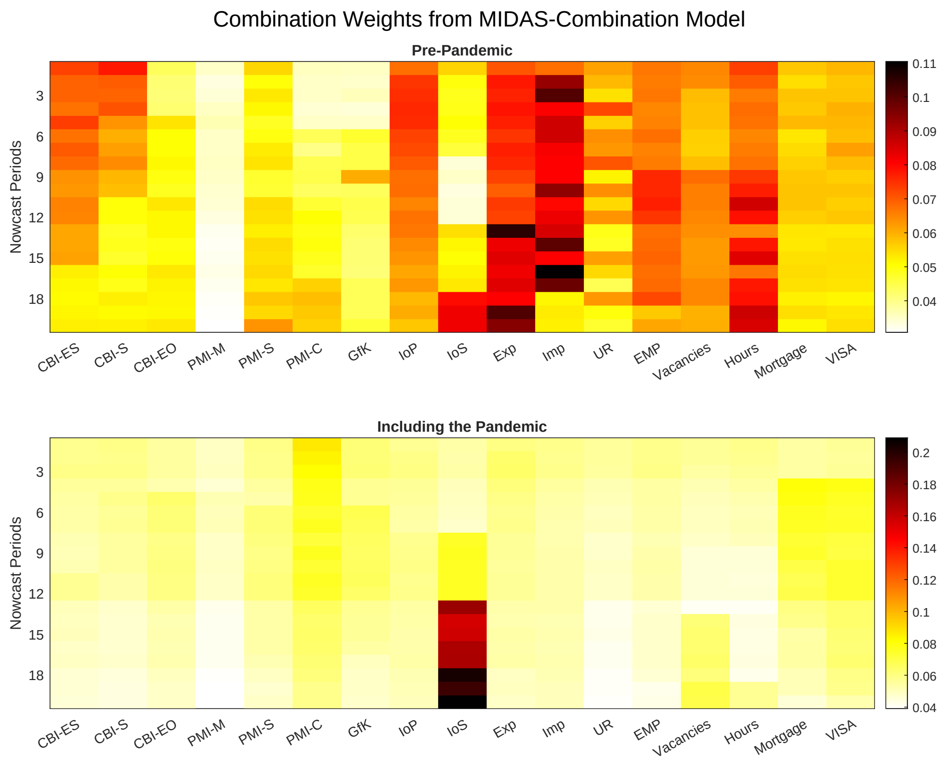


Figure B4: Combination weights of the Combination model of high-frequency variables. Weights pre-pandemic don't change much quarter to quarter, so are averages across quarters until Q4-2019. Each subsequent panel shows the weights for each quarter separately. Vertical axis in each panel represents a nowcast period, according to the pseudo-publication table.

B.5 Further Simulation Results

Table B2: Monte Carlo Simulations: DGP 2

Prior	K	σ_k	$\overline{\sigma}$	RMSFE	RMSE($\mathcal{A}$)	RMSE($\mathcal{A}^C$)	MCC	TPR	FPR	RMSFE	CRPS
Sparse											
HS	30	0.5	0.551	**0.012**	**0.027**	**0.001**	**0.394**	**0.456**	**0.096**	**0.776**	**0.955**
		0.9	0.480	**0.012**	**0.027**	**0.002**	**0.312**	**0.417**	**0.127**	**0.640**	**0.805**
	50	0.5	0.548	**0.011**	**0.030**	**0.004**	**0.292**	**0.439**	**0.129**	**1.027**	**0.910**
		0.9	0.468	**0.019**	**0.040**	**0.012**	**0.231**	**0.445**	**0.213**	**0.885**	**0.908**
	100	0.5	0.531	**0.009**	**0.031**	**0.004**	**0.212**	**0.425**	**0.116**	**0.773**	**0.633**
		0.9	0.434	**0.015**	**0.036**	**0.012**	**0.102**	**0.375**	**0.214**	**0.978**	**0.775**
GAL-SS	30	0.5	0.551	0.36	0.36	0.57	1.51	1.73	1.38	0.76	0.32
		0.9	0.48	0.69	0.70	0.70	1.61	1.49	0.86	0.76	0.29
	50	0.5	0.548	0.31	0.34	0.16	2.15	1.88	0.67	0.61	0.36
		0.9	0.468	0.35	0.49	0.15	2.25	1.44	0.37	0.37	0.22
	100	0.5	0.531	0.29	0.34	0.12	2.93	1.86	0.37	0.73	0.49
		0.9	0.434	0.27	0.50	0.05	5.27	1.55	0.14	0.43	0.29
GIGG	30	0.5	0.551	0.35	0.36	0.52	1.88	1.71	0.51	0.78	0.33
		0.9	0.480	0.60	0.62	0.44	1.95	1.49	0.43	0.76	0.29
	50	0.5	0.548	0.33	0.36	0.19	2.49	1.82	0.32	0.62	0.36
		0.9	0.468	0.30	0.44	0.08	2.78	1.48	0.17	0.36	0.21
	100	0.5	0.531	0.29	0.34	0.12	3.32	1.86	0.22	0.71	0.48
		0.9	0.434	0.23	0.42	0.04	6.28	1.60	0.07	0.42	0.29
Dense											
HS	30	0.5	1.131	**0.028**	**0.032**	**0.008**	**0.127**	**0.516**	**0.381**	**1.870**	**1.127**
		0.9	1.151	**0.035**	**0.040**	**0.017**	**0.084**	**0.511**	**0.422**	**2.347**	**1.429**
	50	0.5	1.417	**0.031**	**0.036**	**0.011**	**0.122**	**0.498**	**0.367**	**1.701**	**1.178**
		0.9	1.406	**0.042**	**0.047**	**0.022**	**0.073**	**0.468**	**0.390**	**1.976**	**1.170**
HS	100	0.5	2.050	**0.038**	**0.043**	**0.015**	**0.071**	**0.443**	**0.367**	**3.280**	**2.009**
		0.9	2.033	**0.044**	**0.051**	**0.019**	**0.045**	**0.402**	**0.355**	**3.028**	**1.995**
GAL-SS	30	0.5	1.131	0.69	0.70	0.49	2.65	1.26	0.77	0.75	0.68
		0.9	1.151	1.05	1.06	0.87	1.92	0.80	0.59	0.74	0.65
	50	0.5	1.417	0.87	0.88	0.53	2.14	1.04	0.65	0.88	0.73
		0.9	1.406	0.99	1.01	0.69	1.86	0.69	0.50	0.89	0.81
	100	0.5	2.050	1.17	1.15	1.25	1.95	0.71	0.49	1.03	0.94
		0.9	2.033	1.16	1.18	0.75	2.33	0.49	0.31	0.81	0.69
GIGG	30	0.5	1.131	0.71	0.72	0.57	2.86	1.17	0.56	0.75	0.70
		0.9	1.151	0.89	0.91	0.74	2.08	0.89	0.65	0.73	0.63
	50	0.5	1.417	0.88	0.88	0.70	1.98	1.03	0.69	0.95	0.78
		0.9	1.406	0.89	0.92	0.60	1.62	0.84	0.70	0.82	0.75
	100	0.5	2.050	1.09	1.09	1.08	1.49	0.92	0.81	0.94	0.84
		0.9	2.033	1.02	1.04	0.80	1.56	0.81	0.72	0.80	0.68

Notes: .