A New Bayesian MIDAS Approach for Flexible and Interpretable Nowcasting*

David Kohns† Galina Potjagailo ‡

Heriot-Watt University, UK Bank of England

September 1, 2022

Preliminary, work in progress.

Abstract

We propose the T-SV-t-BMIDAS model for nowcasting quarterly GDP growth. The model incorporates a long-run time-varying trend (T) and t-distributed stochastic volatility accounting for outliers (SV-t) into a Bayesian multivariate MIDAS. To address over-parameterisation while accounting for group-correlation in mixed frequency data, and to make the model interpretable to the policy maker, we propose a new group-shrinkage prior combined with a sparsification algorithm for variable selection. The prior flexibly accommodates between-group and within-group sparsity and allows to communicate importance of predictors over the data release cycle. We evaluate the model for UK GDP growth nowcasts over the period 1999 to 2021. Our model is competitive prior to the Covid-19 pandemic relative to various benchmark models, while yielding substantial nowcast improvements during the pandemic. First, accounting for a long-run trend and t-distributed stochastic volatility substantially improves forecast performance relative to a simple BMIDAS. Second, the shrinkage prior enhances nowcast performance by inducing group-wise sparsity while enabling the model to flexibly shift between signals. During the Covid-19 pandemic, the model reads stronger signals from indicators for services, which reflected spending shifts related to lockdowns, and less from production surveys. This helps to precisely nowcast the intitial recovery after the shock, and to update the nowcast for the pandemic-related trough sooner. Simulations show competitive performance of the variable selection methodology, with particularly good performance to be expected for highly correlated data as well as dense data-generating-processes.

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 GDP. The large size of the shock caused parameter instabilities in time series methods and has challenged the assumption of a constant variance or slow-moving stochastic volatility in the conditional GDP growth distribution. Additionally, the Covid-19 shock hit economic sectors in a strongly heterogeneous way and standard survey indicators typically employed for nowcasting 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 allow for robust and interpretable nowcasts in presence of large shocks, heterogeneous signals, and regime shifts. This remains relevant given the continuing large economic distortions following the pandemic, Russia's invasion of Ukraine, and the high inflation rates observed across advanced economies.

Since the pandemic, nowcasters and forecasters have pursued different strategies to improve model flexibility, such as down-weighting extreme observations via outlier components in VARs (Lenza and Primiceri, 2022; Carriero et al., 2021) and DFMs (Antolin-Diaz et al., 2021), dropping Covid-19 observations (Schorfheide and Song, 2021), or employing non-parametric non-linear approaches to accommodate extreme observations (Huber et al., 2020). New data sources at a daily or weekly frequency within factor models (Ng, 2021; Baumeister et al., 2021) or in non-parametric predictive regressions (Woloszko, 2020; Kapetanios et al., 2022).

On the other hand, Mixed-data sampling (MIDAS) regressions (Ghysels et al., 2007) have remained under-exploited in nowcasting GDP since the pandemic, despite their competitive nowcast performance in the past (Clements and Galvão, 2009; Foroni et al., 2015; Carriero et al., 2019) and their popularity in policy institutions (Anesti et al., 2017). Unlike factor models, MIDAS models do not rely on co-movement in data but instead include higher-frequency indicators and their lags individually. This can be desirable to exploit heterogeneous signals from the data flexibly during a large shock that affects economic sectors differently. However, multivariate MIDAS models are also prone to various issues: First, such models are highly parameterised, and therefore prone to overfitting without efficient shrinkage (Huber et al., 2019; Frühwirth-Schnatter and Wagner, 2010). This also makes it challenging to extend such models with non-linear and flexible model features. Second, the typically large group-correlation between high-frequency lags can cause nowcasts to randomly load onto certain lags within a group which can hurt both nowcast performance and interpretability of nowcast signals.

In this paper, we propose the Trend-SV-t BMIDAS for flexibly nowcasting GDP before and during the Covid-19 pandemic. We incorporate extreme observations and shifts over time via a time-varying trend and via flexible outlier components into a multivariate MIDAS framework. The model comprises 1) a cyclical component that captures the responses of GDP to a set

of higher frequency indicators, 2) a time varying trend component to capture low-frequency changes in GDP growth and 3) stochastic volatility processes that account for fat tails in the conditional GDP growth distribution. These model features make the model more flexible in dealing with gradual shifts in long-run GDP growth that, if unaccounted for, could blur signals from high-frequency indicators, as well as with large and sudden shifts in the volatility in the data, as observed during the pandemic.

To address the issues prevalent in the multivariate MIDAS regression component, we employ a version a flexible group-shrinkage prior, the GIGG prior (Group Inverse-Gamma Gamma, Boss et al. (2021)). It belongs to the family of global-local priors, but features a three-tiered hierarchy of shrinkage (overall, group-wise, lag-wise). It flexibly detects group-wise sparsity, i.e. it takes the grouping of lags within higher frequency indicators into account, as well as correlation in the shrinkage between group-members, which is inherently high in MIDAS. This avoids the above mixing issues of traditional shrinkage priors. For the latent trend and stochastic volatility components, we employ normal-type priors on the state variances using non-centred state space methods (Frühwirth-Schnatter and Wagner, 2010).

In order to communicate variable importance for the nowcasts from the MIDAS posterior, we propose to combine the GIGG prior with a new sparsification algorithm for group-wise variable selection motivated by Bayesian decision theory. The algorithm finds those 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 unimportant high-frequency indicators to exact zeros, thus allowing to derive inclusion probabilities at each point in time. This allows us to communicate signals drawn upon over time and across groups of predictors, thus enhancing the interpretability of nowcasts.

In an empirical application of the Trend-SVt-BMIDAS model we nowcast UK quarter-on-quarter GDP growth 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 present three main empirical results. First, the posterior estimates of the time-varying components from the Trend-SV-t BMIDAS model conform to economic intuition. We document a gradual decline in the trend component in UK GDP growth since the early 2000s. The Covid-19 shock is identified as a cyclical phenomenon and a sharp increase in GDP growth volatility. The long-run trend in GDP growth remains mostly stable during the pandemic, albeit estimated with higher uncertainty.

Second, the proposed model provides nowcast gains against simpler specifications without time-varying components and against frontier nowcasting models, where our model performs better during the Covid-19 pandemic.we evaluate the proposed model in terms of nowcast precision, comparing the additional benefit of its various components against simpler specifications.

The results show that accounting either for a time-varying trend or for SV with t-errors substantially improves the nowcasts, and that accounting for both provides further gains, particularly pre-pandemic. Intuitively, the inclusion of t-errors helps discount noisy observations for each of the model components. We also find that the Trend-SV-t BMIDAS model is competitive against frontier nowcasting benchmark models such as Combined MIDAS and a DFM, outperforming the benchmarks in late nowcast periods once real activity indicators for the nowcast quarter are released. The relative improvement is particularly pronounced once the Covid-19 shock is included in the sample, where our model can detect the initial economic trough earlier on and nowcast the economic recovery more precisely.

Third, we unpack intuition on the drivers of the nowcast performance via inclusion probabilities of the indicators into the model over the data release cycle. We find that due to the group-shrinkage through the GIGG prior the model draws on a sparse subset of indicators throughout the data release cycle, reading signals from survey indicators during early nowcasts, and then shifting towards signals from a few real activity indicators, particularly the index of services. When including the pandemic, 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. As such, our model is better able to capture the heterogenous nature of the Covid-19 shock.

We also provide sensitivity analysis which shows that our model retains its competitive performance under different MIDAS transformations. We also check nowcast performance during the pandemic when holding MIDAS coefficients fixed at their pre-pandemic levels, as advocated by Schorfheide and Song (2021). We find that this markedly decreases nowcast precision, which indicates that the flexible features of our model help in nowcasting the large shock.

Our finding of a strong nowcasting performance of a group-sparse model contributes to the general debate on the "illusion of sparsity" in macroeconomic forecasting (Giannone et al., 2021; Fava and Lopes, 2021; Kohns and Bhattacharjee, 2020). We find that grouping structure is important to account for in macro applications when analyzing sparsity patterns and that sparse models can outperform dense specifications in the context of our model, particularly in presence of a heterogeneous shock.

Lastly, we provide a simulation study that investigates how well the proposed combination of GIGG prior with sparisification generalise to a host of different MIDAS settings. We compare the prior to two frontier priors and show that our approach yields competitive performance with particularly good performance to be expected in high correlated data as well as dense data-generating-processes.

The remainder of the paper is structured as follows. Section 2 presents the Trend-SV-t Bayesian MIDAS approach and the proposed flexible priors. Section 3 outlines the data set

and setup of our nowcast application. Section 4 present the results on the in-sample model features, nowcast performance and variable inclusion probabilities. Section 5 investigates how well the proposed prior and sparsification algorithm generalise to a host of sparse and dense data-generating processes. Section 6 concludes. More details on the methodology and additional results can be found in Appendix 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 three main features of the model: cyclical MIDAS, trend and stochastic volatility components. Then, we present the group-shrinkage prior which we argue is well suited to the MIDAS estimation due to its ability to take the mixed-frequency grouping and correlation structure into account, alongside its aggressive shrinkage which facilitates variable selection. Lastly, we detail our decision theory inspired variable selection approach that allows to communicate which high-frequency variable groups have the highest impact for the nowcasts.

2.1 The model

The proposed Trend-SV-t BMIDAS model for nowcasting quarterly GDP growth is a flexible unobserved components model that includes three main features: (1) a multivariate MIDAS regression component that captures cyclical macroeconomic information from a range of higher-frequency (monthly) indicators, (2) a trend component that captures latent slow-moving changes in GDP growth, 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 cyclical component $Z_t^{(m)} = (z_{1,t}^{(m)}, \cdots, z_{K,t}^{(m)})'$ 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 parameters $\theta = (\theta_1', \cdots, \theta_K')'$ measure the response to changes in the K macroeconomic indicators, each with L_k monthly lags[1]. θ_k features $p_k + 1$ parameters that act as a linking function between these higher frequency lags 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 spirit of Almon (1965) that has been recently popularised for high dimensional mixed frequency forecasting applications by Mogliani and Simoni (2021)[2]. Parsimony is induced in linear Almon-MIDAS by assuming a $p_k << L_k$ polynomial process of the coefficients across high-frequency lags, 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 save degrees of freedom, the cyclical component can still suffer from multicollinearity due to the high degree of serial correlation present in the mixed frequency lags $Z_t^{(m)}$. This will be directly addressed with our shrinkage prior in section 2.2.2.

The trend component in (2) dictates driftless random walk behaviour which captures the low frequency changes in GDP growth. Driftless trends in GDP growth, with state variances tightly controlled by priors to enforce slow-moving change, have been shown to markedly improve nowcasts in applications to the US (Antolin-Diaz et al., 2017, 2021). This reflects the empirical finding of changes in long-run GDP growth (Kim and Nelson, 1999; McConnell and Perez-Quiros, 2000; Jurado et al., 2015) that indicate slow moving shifts in general economic conditions that are otherwise not well captured by high-frequency information.[3]

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. Volatility clustering is an empirical regularity in aggregate macroeconomic time-series data (Stock and Watson, 2007), and has been shown to significantly improve density and point forecasts in many forecasting studies (Koop and Onorante, 2019; Huber et al., 2019) and mixed frequency models (Carriero et al., 2015; Antolin-Diaz et al., 2017). 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.[4] The more commonly employed centred SV

[1] Lags 0 to 2 correspond to leads, while 3-5, correspond to lags in relation to the quarterly time-steps, t.

[2] 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.

[3] 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.

[4] Since w_g and w_h, driectly appear in the observation and state equation, which control the overall state

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-lividness 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.

While the role of long-run trend dynamics and stochastic volatilities with fait tails for nowcasting and modelling GDP growth has recently been documented for VARs and mixed-frequency dynamic factor models Lenza and Primiceri (2022); Antolin-Diaz et al. (2017, 2021), to our knowledge we are the first to incorporate time-varying trends and stochastic volatility processes into a MIDAS setting.

2.2 Bayesian Setup

To regularise parameter estimation variance, we make use of Bayesian priors. Regularisation is key since both the latent states as well as the multivariate MIDAS structure are highly parametrised. We first present the general prior framework that gives rise to the conditional posteriors, before detailing in 2.2.2 the form of the proposed GIGG prior and give further intuition about its behaviour.

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

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).

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:

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

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. 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 $\mathcal{O}(T^2 \times (\sum_{k=1}^{K} p_k + 1))$.[5]

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 nowcasting 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 as a whole has for the nowcast target.

We therefore use a three-tiered shrinkage prior that addresses these issues jointly, as it

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

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 shrink jointly to zero, but the relevant groups feature a high degree of correlation among individual lags, instead of a heterogeneous shrinkage within group. In this, setting small a_k and relatively large b_k can be a good fit 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 $a_k = 1/T, b_k = 0.5 \; \forall k$.

One the other hand, in situations in which prior knowledge exists that only selected lags are important, a relatively small b_k may help the prior to shrink individual lags within a group more idiosyncratically toward zero. As a robustness check, we report in the appendix (B.3) nowcast cast results for different hyper-parameter choices.

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 1[6] and for other combinations of the hyper-parameters, follows a correlated normal beta prime distribution akin to Armagan et al. (2013), $\gamma_k^2\varphi_{k,j}^2 \sim \beta'(a_k, b_k)$. 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).[7]. 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 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

[6] 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 1.

[7] 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)

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 hyperparameter 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 the nowcasts. Thresholding those lag groups with little effect on y_t to exactly zero, i.e. essentially eliminating them from the model, makes it easier to interpret and communicate model outcomes and 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); Woo and Owen (2019).[8] 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 nowcasts:

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

where $\tilde{\boldsymbol{Y}}$ refers to a realisation from the posterior predictive distribution $p(\tilde{\boldsymbol{Y}}|\boldsymbol{y}) = \int p(\tilde{\boldsymbol{Y}}|\boldsymbol{y}, \boldsymbol{Z}^{(\boldsymbol{m})}, \theta, \bullet) p(\theta|\boldsymbol{y}, \boldsymbol{Z}^{(\boldsymbol{m})}, \bullet) d\theta$, and $|| \bullet ||_p$ refers to the ℓ_p-norm.[9] 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

[8] 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.

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

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.[10] 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 will be used to report inclusion probabilities that inform on the relative impact of an indicator for nowcasting. 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 Nowcast Setup

For our application, we nowcast real quarter-on-quarter GDP growth of the UK 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 nowcasting 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

[10] 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)

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.[11]

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.[12] We use final vintages of the data, downloaded in December 2021. While revisions of UK monthly real activity indicators and quarterly GDP can be substantial, accounting for them accurately in the nowcast set-up also poses computational and methodological challenges.[13] Since our focus here lies in understanding the proposed model, we use pseudo real time data as outlined below, and we leave an account for revisions for future research.

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).

As is common in the MIDAS nowcasting literature, we start the nowcast exercise for each quarter anew, where we start predicting with all available information on the first of the month prior to the reference quarter (forecasts), and simulate incoming data throughout the reference quarter (nowcasts), until GDP is finally observed six weeks after the reference quarter (backcasts).[14] For each new data release over the data release cycle, we generate nowcasts from the

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

[12] 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.

[13] 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.

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

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

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

Notes: "Timing" refers to typical data release times as of December 2021, abstaining from changes in the publication calendar over the sample period. "Release" refers to the data series updated at a given nowcast, 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).

predictive distribution $p(y_{t+1}|\Omega_T^v)$, where $(v = 1, \cdots, 20)$ refers to the nowcast periods within the stylised data release calendar (Table 1) and Ω_T^v represents the real-time information set that expands with each data release. Since the MIDAS framework belongs to the class of reduced system mixed frequency models (Bańbura et al., 2013), each information set Ω_T^v results in a different model (depending on which data are observable to the nowcaster). [15] 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

[15] This is in contrast to full system mixed frequency methods (Bańbura et al., 2013) which interpolate missing low- and high-frequency information via filtering methods.

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 nowcasting 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 4.2, evaluates nowcast performance of the Trend-SV-t-BMIDAS model against a simpler BMIDAS without additional model features, and then against commonly employed benchmark models. Section 4.3 then unpacks the signals behind the nowcast estimates via inclusion probabilities derived by the decision theoretic variable selection method.

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.

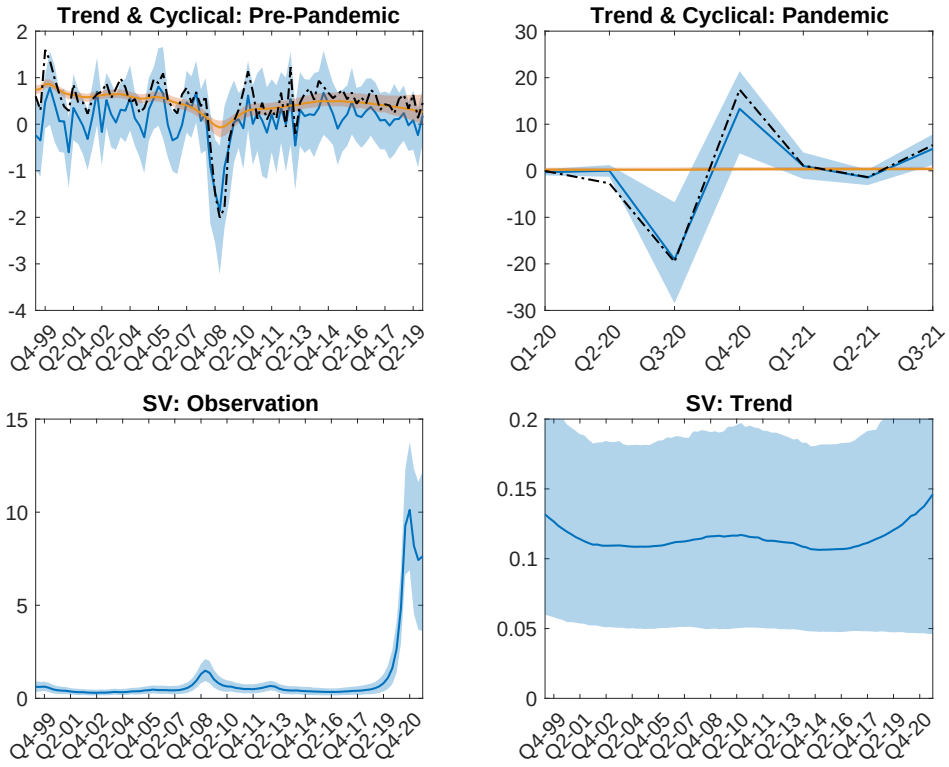


Figure 1: Posterior estimates for trend, cyclical component and stochastic volatilities.

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.

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

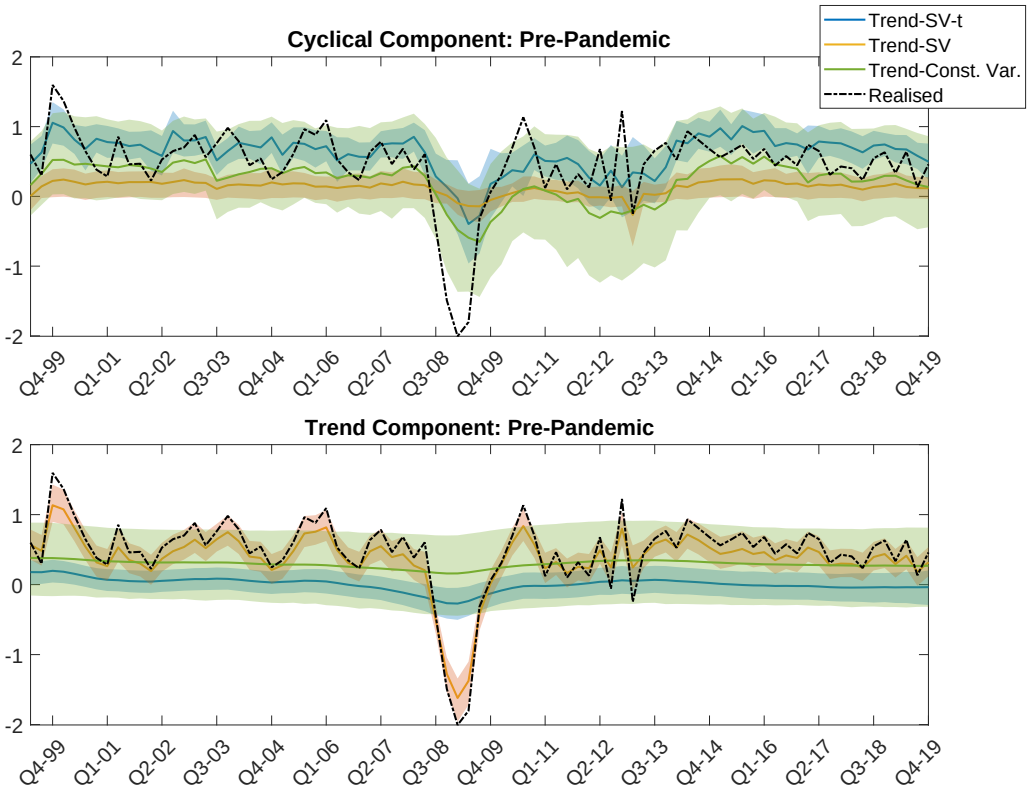


Figure 2: Posterior trend estimates from different BMIDAS specifications, pre-pandemic.

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.

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 no t-distr. errors (orange), as well as for constant variance only (green).[16] 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,

[16] 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.

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 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 for nowcasting performance. Then, we evaluate the preferred Trend-SV-t model against a range of benchmarks, with a focus on nowcasting performance during the pandemic.

4.2.1 Role of model features for nowcasting

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 volatilies, 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 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

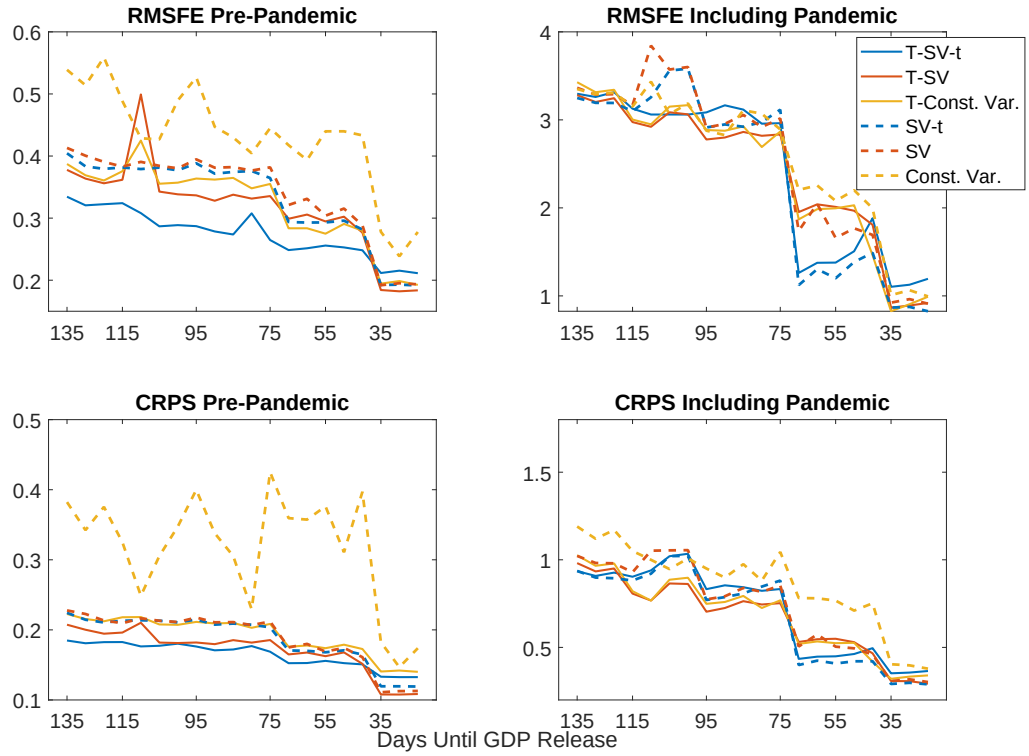


Figure 3: Nowcast performance for different BMIDAS model components.

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.

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 around nowcasts 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 nowcast 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 nowcast 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 Nowcast evaluation against benchmark models

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-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.[17]

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 nowcast studies (e.g.

[17] 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.

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.

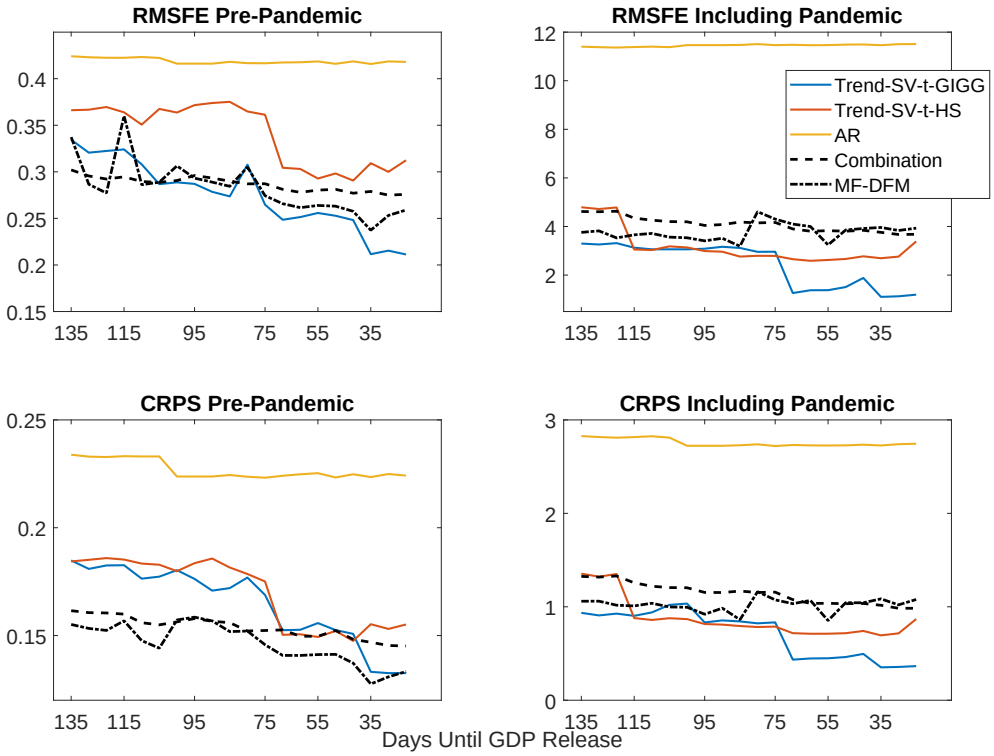


Figure 4: Nowcast performance of Trend-SV-t-BMIDAS with GIGG compared to benchmarks.
Notes: Absolute RMSFE and CRPS over nowcast periods (days before GDP release on x-axis) for Trend-SV-t-BMIDAS with GIGG prior (blue lines), Trend-SV-t-BMIDAS with HS prior (orange), Combined U-BMIDAS (black dashed), 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.

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

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)

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 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.[18]

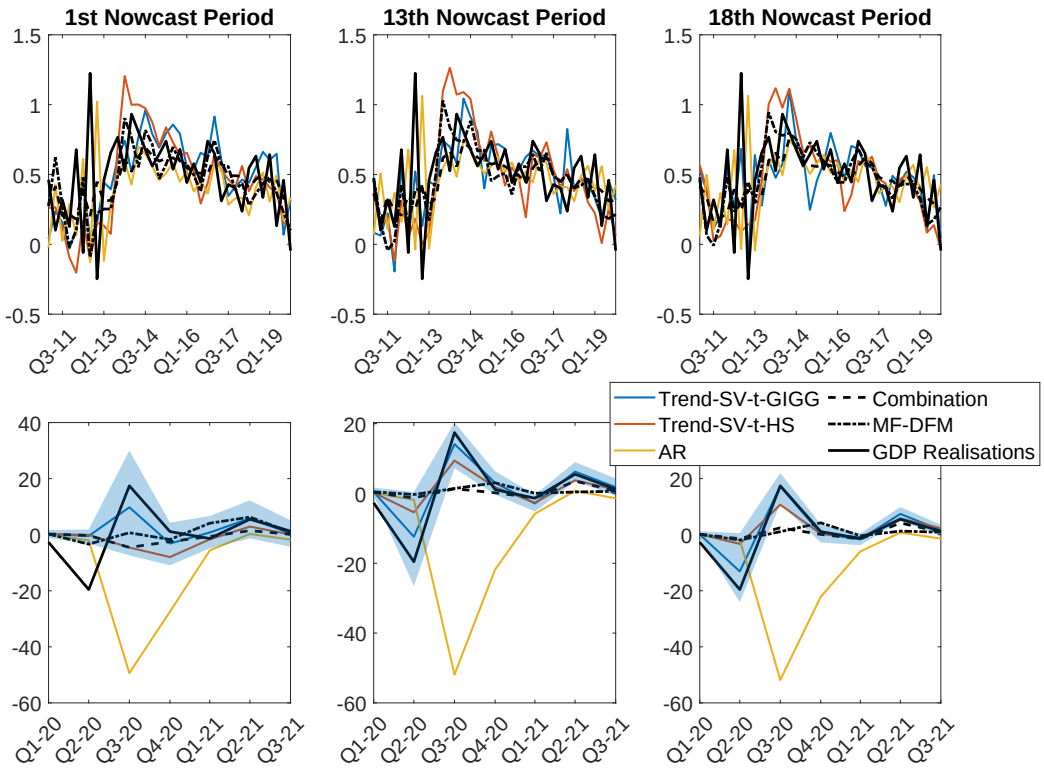


Figure 5: Posterior mean and density nowcasts, selected nowcast periods.

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.

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

[18] 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.

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.

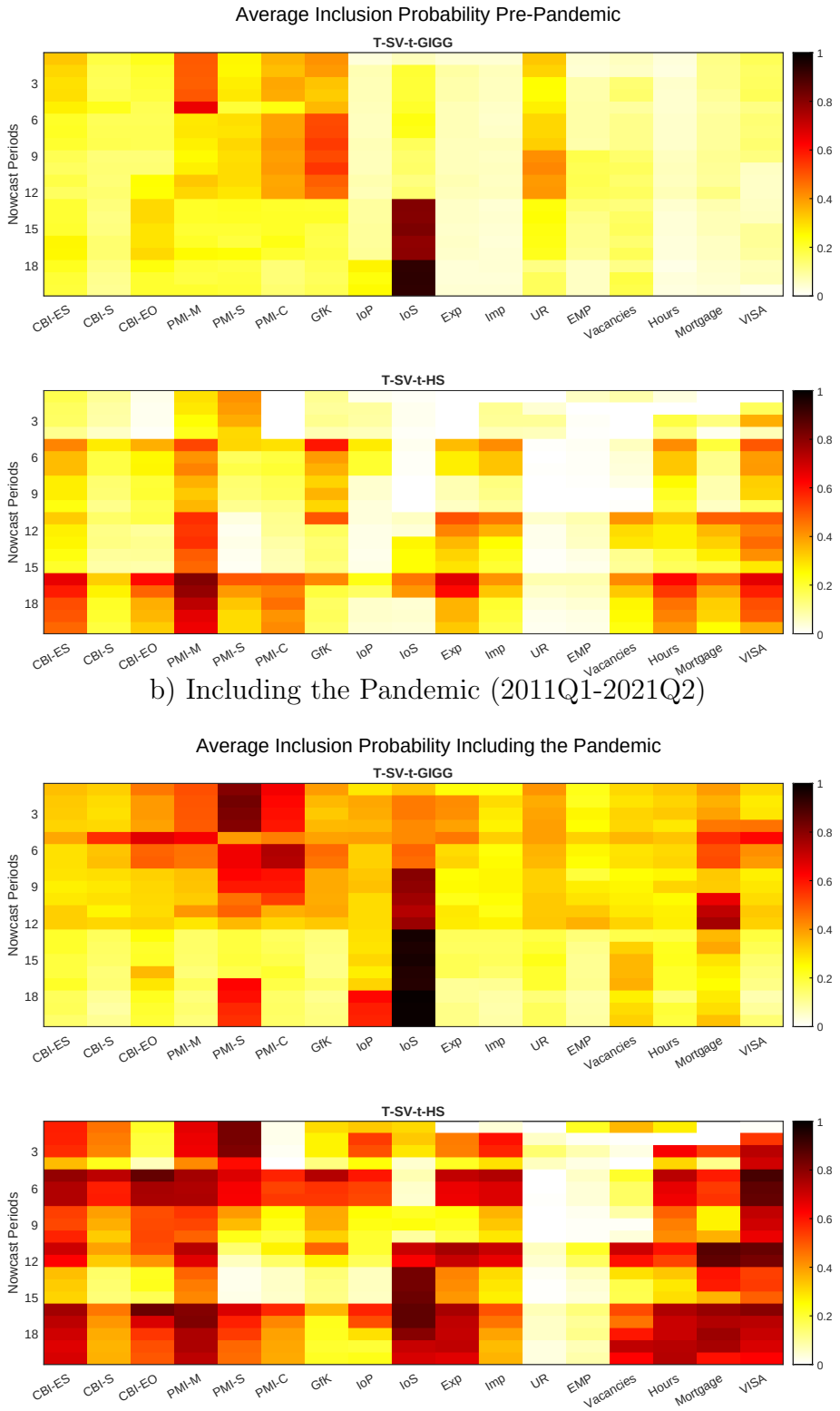
4.3 Interpreting Nowcasts via Variable Inclusion Probabilities

The group-variable selection achieved via the sparsification algorithm (9) equips us with an intuitive way to communicate nowcast 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 nowcast 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 nowcast studies who find 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

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).

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 forecast 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. Nowcast evaluation results in comparison to the baseline specification are presented in Figure (7).

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

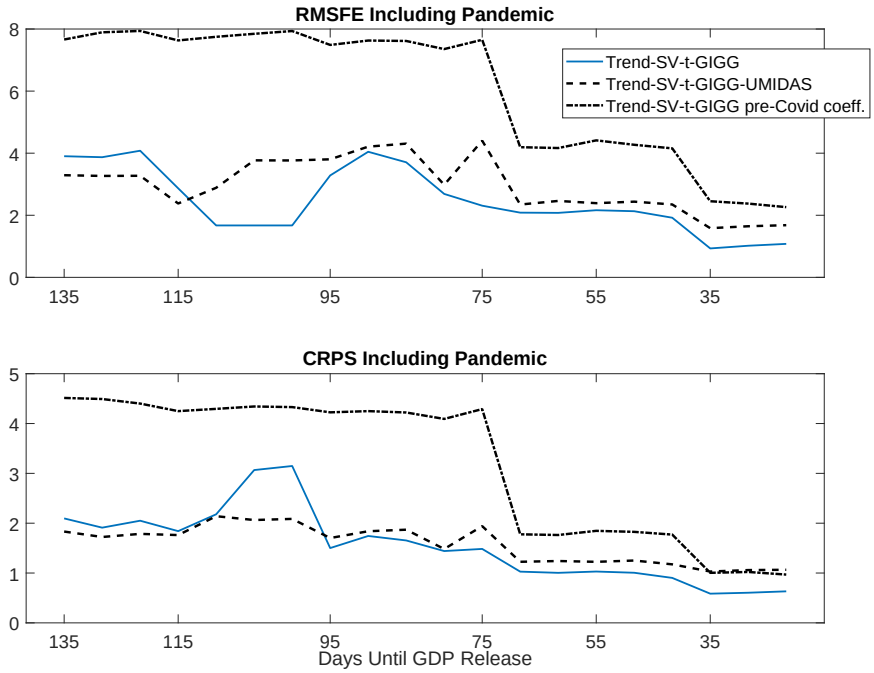


Figure 7: Nowcast performance of alternative specifications of the Trend-SV-t-BMIDAS.
Notes: The Trend-SV-t-GIGG-UMIDAS uses U-MIDAS sampled data (Foroni et al., 2015) instead of Almon restricted lags. The Trend-SV-t-GIGG pre-Cov.coeff nowcasts Q1-2020 onwards with coefficients based on the sample that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

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 Simulation Study

In the nowcasting exercise using the GIGG prior, we find evidence for a sparse set of groups that drive the cyclical component. Since this is in opposition to the existing canon in the macro forecasting literature which finds that dense components usually drive forecasts of aggregate economic time-series, we investigate in this section parameter estimation, variable selection and forecasting precision via a simulation study. In particular, we are interested how our model's performance generalises to a variety of different MIDAS data-generating-processes (DGPs) where we change correlation structures, degrees of sparsity and MIDAS lag profiles. And to abstract from any estimation issues stemming from time stochastic components, we consider in this section only DGPs without Trend and SV-t components. This also serves to facilitate comparability to similar simulation studies such as in Mogliani and Simoni (2021). We generate data from the following model:

$$\begin{aligned} y_t &= \alpha + \sum_{k=1}^{K} \beta_k \sum_{j=0}^{p_k+1} \theta_j z_{j,k,t-h}^{(m)} + \epsilon_t^y,\ \epsilon_t \sim N(0, \sigma^2) \\ x_{k,t-h}^{(m)} &= \mu + \rho x_{k,t-h-1}^{(m)} + \epsilon_{k,t},\ k = 1, \cdots, K, \end{aligned} \tag{12}$$

where $x_{k,t}^{(m)}$ are the untransformed high-frequency data ($X_t^{(m)} = (x_{1,t}^{(m)}, \cdots, x_{K,t}^{(m)})$), and $z_{k,t}^{(m)}$ are the MIDAS transformed high-frequency lags. This notation distinguishes between the MIDAS lag profile, modelled via normalised weights θ_j, and the impact it has on y_t, β_k. By including zero entries in β_k, we are able to control the group-wise degree of sparsity. High-frequency data are considered in monthly frequency, $m = 3$, and we consider $L_k = 24$ monthly lags, to make the simulation study comparable to Mogliani and Simoni (2021)[19]. The correlation structure of the high-frequency data, $X_t^{(m)}$, are determined by $\epsilon_t^K = (\epsilon_{1,t}, \cdots, \epsilon_{K,t}) \sim N(0, \Sigma_K)$, where Σ_K has elements $\sigma_k^{|k-k'|}$.

We vary the DGP along 4 different dimensions: 1) the number of groups, K, 2) the sparsity in β, 3) cross-sectional high-frequency correlation σ_k, and 4) the MIDAS lag profile in θ. We consider relatively small to large dimensions, $K = \{30, 50, 100\}$. For β, we generate from a sparse regime $\beta = (0, b_1, b_2, 0, b_3, b_4, 0, 0, b_5, \mathbf{0})$, and a dense regime where we populate randomly

[19] We tested the simulations with $L_k = 6$ as in the empirical application. The results are comparable and available upon request

70% of the entries with non-zero coefficients. For both sparsity regimes, we generate $b_k \sim N(0,1)$ such that both small and large coefficients are present in the DGPs. To test the prior in both low and high correlation data environments, we set $\sigma_k \in \{0.5, 0.9\}$. Finally, for the weight profiles θ, we follow Andreou et al. (2010) and Mogliani and Simoni (2021) by considering normalised weights which follow a steep, moderately steep and flat decay across the 24 high frequency lags, visualised in Figure 8.[20] The profiles also serve to test how well the prior adapts to varying degrees of within-group sparsity. The steep profile sets coefficients for most lags to 0, while the flat profile's coefficients are entirely non-zero. We generate 100 simulated data set for each DGP variant.

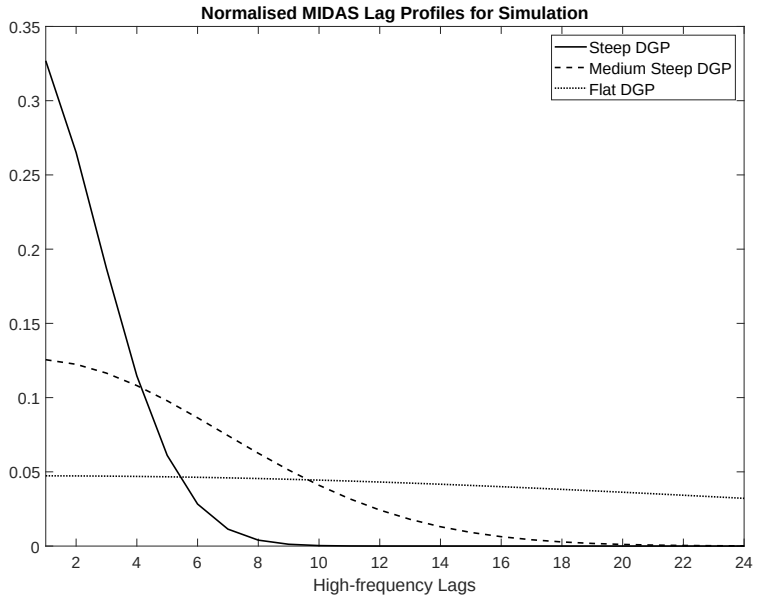


Figure 8: MIDAS lag profiles for a steeply, medium steeply, and flatly decaying high-frequency lag profiles over 24 for lags. These are generated from a normalised non-linear Almon function as in Andreou et al. (2010).

We specify the GIGG prior, as in the empirical application with the hyper-parameters set to $a_k = 1/T, b_k = 0.5, \ \forall k$, and apply the sparsification algorithm in 9 at each Gibbs draw. To put the performance of the GIGG prior into perspective, we consider the non-grouped horseshoe prior with group-sparsification as in the empirical application, as well as the recently proposed group prior of Mogliani and Simoni (2021), the group-adaptive-lasso prior with spike-and-slab sparsification (GAL-SS). Mogliani and Simoni (2021) show that the GAL-SS compares favourably to competing lasso and group spike-and-slab shrinkage priors. The GAL-SS, similar to the GIGG prior, has an adaptive group-level shrinkage scale, but conducts variable selection via group-spike-and-slab approach, rather than a decision theoretic approach and does not

[20] The weights are generated from a non-linear Almon function. See for more details Andreou et al. (2010).

feature an individual covariate level shrinkage scale to model the degree of correlation between the mixed frequency parameters (see Mogliani and Simoni (2021) for further details). Priors on the error variance, σ, for all models are assumed to be uninformative.

5.1 Evaluation

To evaluate the priors' performance we measure the average root-mean-squared-error (RMSE) of the MIDAS coefficients:

$$\frac{1}{M}\sqrt{||\hat{\theta} - \theta||_2^2}, \tag{13}$$

where M is the number of simulations, set to 100, and $\hat{\theta}$ is the posterior mean based on the retained MCMC draws. To support the information about zero (i.e. "inactive") vis-a-vis non-zero ("active") components of each DGP, we report RMSE for each set also separately as $RMSE(\mathcal{A}^C)$ and $RMSE(A)$, respectively. To gauge nowcasting performance, we generate predictions for the last observation of each sample, and treat the information set as fully available. These are evaluated via the RMSFE based on the the mean of the posterior predictive distribution, as as the CRPS, following 11. Predictions are generated, as in the empirical application via Monte Carlo integration.

Lastly, to measure the precision on variable selection, we compute Matthew's correlation coefficient (MCC) as well as the true positive rate (TPR) and false positive rate (FPR). The the MCC gives an overall measure of goodness of variable selection based on the true and false positives and negatives and the TPR and FPR indicate how well the model selects the true set and avoids the false set, respectively. These are calculated as:

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{14}$$

$$TPR = \frac{TP}{TP + FN} \tag{15}$$

$$FPR = \frac{FP}{FP + TN} \tag{16}$$

where T stands for "True", F stands for "False", N stands for "Negative", and P stands for "Positive". The MCC is between -1 (worst) and +1 (best), while the TPR and FPR rate will range between 0 and 1. Higher TPR and lower FPR are better, ceteris paribus.

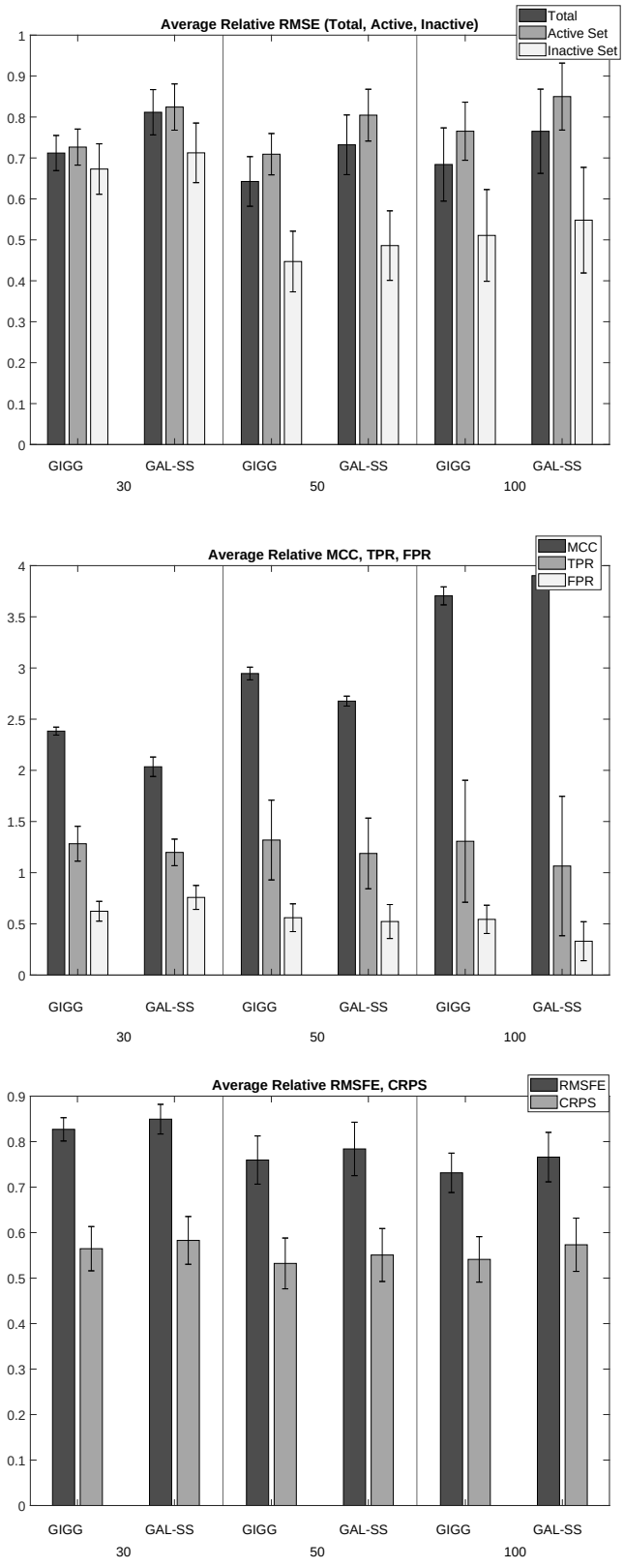


Figure 9: Average relative performance for a) the parameter estimation precision $(RMSE, RMSE(\mathcal{A}), RMSE(\mathcal{A}^C))$, b) variable selection (MCC,TPR,FPR) as well as nowcasting (RMSFE,CRPS). Each of these metrics are relative to those of the horseshoe prior benchmark. The horizontal axis of each panel is grouped into the two group-priors and by the covariate dimension $K = \{30, 50, 100\}$.

Table 3: Monte Carlo Simulations: Sparse Coefficient Vector

Prior	K	σ_k	$\overline{\sigma}$	RMSFE	RMSE($\mathcal{A}$)	RMSE($\mathcal{A}^C$)	MCC	TPR	FPR	RMSFE	CRPS
DGP 1: Steep MIDAS Coefficients											
HS	30	0.5	0.754	**0.017**	**0.041**	**0.003**	**0.327**	**0.539**	**0.197**	**1.379**	**0.984**
		0.9	0.668	**0.022**	**0.053**	**0.005**	**0.288**	**0.558**	**0.239**	**1.065**	**0.873**
	50	0.5	0.739	**0.016**	**0.046**	**0.007**	**0.295**	**0.59**	**0.203**	**1.123**	**0.761**
		0.9	0.608	**0.021**	**0.05**	**0.011**	**0.248**	**0.475**	**0.202**	**0.977**	**1.136**
	100	0.5	0.688	**0.012**	**0.046**	**0.005**	**0.281**	**0.588**	**0.139**	**1.046**	**0.729**
		0.9	0.615	**0.02**	**0.068**	**0.012**	**0.127**	**0.441**	**0.217**	**0.899**	**0.789**
GAL-SS	30	0.5	0.754	0.7	0.72	0.36	1.71	1.33	0.62	0.82	0.64
		0.9	0.668	0.78	0.78	0.48	1.85	1.14	0.41	1.02	0.59
	50	0.5	0.739	0.57	0.63	0.1	2.03	1.29	0.39	0.63	0.57
		0.9	0.608	0.82	0.99	0.46	1.82	1.14	0.35	0.71	0.34
	100	0.5	0.688	0.53	0.6	0.12	2.14	1.25	0.29	0.82	0.62
		0.9	0.615	0.53	0.68	0.09	4.2	1.31	0.14	0.73	0.44
GIGG	30	0.5	0.754	0.67	0.68	0.64	1.85	1.39	0.55	0.81	0.64
		0.9	0.668	0.66	0.66	0.53	1.91	1.23	0.47	0.97	0.57
	50	0.5	0.739	0.55	0.60	0.23	2.02	1.34	0.44	0.61	0.5
		0.9	0.608	0.55	0.7	0.22	2.14	1.36	0.40	0.69	0.33
	100	0.5	0.688	0.53	0.58	0.23	1.94	1.32	0.45	0.82	0.63
		0.9	0.615	0.47	0.6	0.10	4.22	1.47	0.22	0.75	0.43
DGP 3: Flat MIDAS Coefficients											
HS	30	0.5	0.388	**0.011**	**0.025**	**0.002**	**0.246**	**0.349**	**0.122**	**0.853**	**0.246**
		0.9	0.428	**0.015**	**0.028**	**0.007**	**0.164**	**0.381**	**0.220**	**0.651**	**0.812**
	50	0.5	0.361	**0.011**	**0.028**	**0.005**	**0.150**	**0.290**	**0.142**	**0.674**	**0.801**
		0.9	0.371	**0.017**	**0.03**	**0.012**	**0.105**	**0.3**	**0.195**	**1.004**	**0.919**
	100	0.5	0.35	**0.009**	**0.03**	**0.005**	**0.087**	**0.27**	**0.134**	**0.821**	**0.691**
		0.9	0.373	**0.016**	**0.032**	**0.014**	**0.057**	**0.322**	**0.235**	**1.058**	**0.749**
GAL-SS	30	0.5	0.388	0.79	0.80	1.26	1.89	1.94	1.39	0.83	0.43
		0.9	0.428	0.92	1	0.67	1.98	1.21	0.65	0.96	0.40
	50	0.5	0.361	0.60	0.71	0.34	3.12	2.35	0.84	0.98	0.45
		0.9	0.371	0.50	0.79	0.23	3.74	1.71	0.49	0.59	0.36
	100	0.5	0.350	0.58	0.66	0.41	4.72	2.38	0.62	0.59	0.38
		0.9	0.373	0.42	0.86	0.14	6.43	1.39	0.21	0.51	0.39
GIGG	30	0.5	0.388	0.76	0.76	1.22	2.27	1.96	0.89	0.77	0.41
		0.9	0.428	0.80	0.87	0.70	2.37	1.40	0.63	0.89	0.38
	50	0.5	0.361	0.57	0.68	0.33	3.61	2.31	0.56	0.94	0.41
		0.9	0.371	0.49	0.72	0.24	4.48	1.82	0.39	0.6	0.36
	100	0.5	0.35	0.54	0.64	0.30	5.52	2.36	0.43	0.62	0.39
		0.9	0.373	0.37	0.79	0.12	7.29	1.57	0.22	0.52	0.4

Notes: The table shows in bold the level of each of the evaluation metrics, while those of the GAL-SS and GIGG are in relative terms to the corresponding entry of the horseshoe model. $\overline{\sigma}$ is the Monte Carlo error of the DGP. These results refer to sparse DGPs only.

5.2 Simulation Results

Tables 3, 4, and B2 in the appendix give the average evaluation metrics over the Monte Carlo simulations. [21] Table 3 focuses on the sparse and table 4 on the dense DGPs. We show the results of the horseshoe prior (bold) in levels as the benchmark, to which the grouped priors are

[21] We leave the medium steep results to the appendix as they are generally in between the steep and flat DGP results

shown in relative terms. To facilitate the discussion, values lower than 1 for RMSE (incl. active and inactive), FPR, RMSFE and CRPS indicate outperformance of the group prior relative to the horseshoe, while values higher than 1 for MCC and the TPR entail outperformance. And to summarise the results, we plot in figure 9 the average relative evaluation metrics which is pooled across all but the coefficient dimensions. We will first review the general findings based on the horseshoe prior and then compare that to the performance of the GIGG and GAL-SS.

In line with previous studies (e.g., Giannone et al. (2021); Huber et al. (2019); Bitto and Frühwirth-Schnatter (2019)), we find an increase in parameter estimation error and a decrease in variable selection accuracy as a function of in cross-sectional correlation, denseness of the DGP as well as the covariate size for the horseshoe benchmark.

The effect of larger correlation comes at no surprise, as it becomes increasingly hard for shrinkage priors to distinguish between signals and noise with higher correlated designs, leading to increases in false selection of inactive groups (see the FPR columns of both tables). This causes a general decrease in the quality of variable selection (column MCC) and an increase in RMSE, with with RMSE for the inactive set (column $RMSE(\mathcal{A}^C)$) increasing more than for the active set (column $RMSE(\mathcal{A})$). It should be noted, though, that the level of RMSE for the active one is always larger than for the inactive one, independent of the DGP. This is a common finding for shrinkage priors (Zhang et al., 2022; Cross et al., 2020; Kohns and Bhattacharjee, 2020), which similar to frequentist lasso methods, retain larger non-zero bias in finite samples for active coefficients than for inactive ones.

In line with this, the horseshoe prior also struggles more with dense DGPs (first panel, table 4) than with sparse DGPs, showing higher RMSE for active, but also inactive sets, along with much lower MCC. This is driven by larger FPR. And since the TPR is also higher with dense DGPs, this suggests that the horseshoe prior over-selects variable groups.

The response to increasing the dimension of the covariate set, K, has a differing effect, most notably depending on the degree of sparsity. While table 3 shows either stable RMSE (flat DGP, 4th panel) or a decrease in RMSE (steep DGP, 1st panel), the dense DGPs in table 4 show clear increases in overall estimation error (flat and steep). Similar findings are reported Mogliani and Simoni (2021) (and references therein) and is related to the fact that the extra inactive covariate groups are less and less correlated with increasing K by design, for which estimation error is lower. Thus total RMSE may even decrease in sparse settings with larger K. Similarly, variable selection suffers less in sparse than in dense DGPs as judged by the MCC.

In terms of nowcasting (columns RMSFE and CRPS), the worse parameter precision and variable selection in dense compared to sparse DGPs result in comparatively much larger RMSE and CRPS. And similar to Mogliani and Simoni (2021), we find that correlation tends to not adversely affect nowcasting performance, while the effect of larger K again depends on the degree of sparsity. This makes intuitive sense as nowcaststs can still be accurate when selecting

wrong groups, as long as the signal for y_t is highly correlated among the covariates.

Finally, when comparing flat to steep DGPs, we find that RMSE is generally lower for the former, despite variable selection (see MCC) being worse. This is likely due to the the flat coefficients already being closer to zero compared to steep DGPs such that falsely eliminating active groups has relatively less effect on overall RMSE. In line with this, the true positive rate is far lower for flat than for the steep DGPs. A complicating factor for the flat DGP in general is that the end-point restrictions falsely force the estimated profiles to end smoothly at zero, and hence, lower TPR compared to steep DGPs, for which the tail restrictions are a good fit, are somewhat expected.

Compared to the horseshoe prior, the group priors improve general performance considerably, with relative RMSE all below 1, MCC larger than 1 and nowcast metrics below 1. In terms of parameter precision, errors for the inactive sets are greatly reduced, while better variable selection is mostly driven by a better FPR, particularly in higher dimensions. Hence, the horseshoe over-selects compared to the grouped priors, which was indicative of the horseshoe already from the nowcast application. It is noteworthy that the better variable selection has greater impact on uncertainty quantification than point forecasts with the CRPS being relatively lower than the RMSE compared to the horseshoe.

Comparing the GIGG to the GAL-SS, figure 9 highlights that the GIGG almost always outperforms the GAL-SS, with the differences remaining remarkably constant across evaluation metrics and covariate sizes. The patterns that emerge are that 1) the GIGG shows relatively better performance in the active coefficient sets than in the inactive ones, 2) GIGG has overall better variable selection as gauged by MCC in $K = \{30, 50\}$, and 3) slightly better nowcasting performance. With respect to variable selection, we find that outperformance in MCC is driven by higher TPR, albeit the incurring slightly larger FPR. Hence, the GAL-SS remains more conservative on average in variable selection compared to the GIGG. These general trends, though, mask some interesting dynamics across individual DGPs.

Table 3 shows that the GIGG prior tends to outperform the GAL-SS in terms of parameter precision when correlation is high, such as the sparse, steep (3, third panel) and sparse, flat DGPs (3, sixth panel), as well as in dense DGPs more generally (4). For both these sets of DGPs, the lower RMSE are driven by comparatively larger decreases in errors for the active coefficients than in the inactive sets[22]. This is likely driven by two interrelated explanations. On the one hand, spike-and-slab priors discretise the parameter space and thus can suffer from worse mixing than continuous shrinkage priors with increases in correlation (Barbieri and Berger, 2004; Piironen et al., 2020), and on the other, the GIGG benefits from the extra covariate level scale which helps in adapting to within group sparsity observed for the steep

[22] For flat DGPs generally, GIGG tends to consistently outperform the GAL-SS, albeit with smaller margins than for the steep DGPs.

DGP in particular. For the dense DGPs in which the active set is relatively more important for overall RMSE, we therefore see much improved performance. As soon as correlation is low, though, the margins between the GAL-SS and GIGG tend to decrease.

Variable selection is very much in line with this narrative. The GAL-SS shows similar MCC for low correlation, sparse DGPs, but falls off compared to the GIGG on all other settings, driven mostly by lower TPR, albbeit an often also lower FPR, particularly in high correlation DGPs.

Forecasting, is often either very similar or slightly better for the GIGG, although the differences are smaller than for parameter precision and variable selection. This is particularly true in the higher correlation settings, where selecting the wrong subgroup is less important. That being said, in large dimensions, larger differences in forecasts appear in dense settings which is likely explained by the larger relative importance of parameter errors in the active set, which is generally lower for the GIGG.

All in all, the results from the simulation exercise confirm the intuition gathered in the exposition of the GIGG prior and the nowcast application. For one, we find the GIGG, aided by more flexible within-group shrinkage, provides smallest estimation error when within-group sparsity is present and provides a large improvement over the horseshoe and even improves on the GAL-SS group-shrinkage prior when correlation is high. And for the other, the GIGG improves over both priors in dense DGPs, for which the accuracy of the active set matter comparatively more. This is also likely facilitated by better mixing offered from the continuous nature of the shrinkage compared to the GAL-SS. The simulations have also shown that the sparsification algorithm provides variable selection that is on par with the GAL-SS or considerably outperforms in the dense and higher correlation DGPs. Hence, these results give further credence to the variable selection results found in the empirical application.

6 Conclusion

In this paper, we have proposed a new Bayesian MIDAS framework, the T-SV-t-BMIDAS model, for nowcasting quarterly GDP growth, combined with a flexible group prior and variable selection algorithm 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 fashion, and works well also in more tranquil times, particularly at a later stage of the data release cycle, when it flexibly draws signals from "hard" indicators rather then survey data only.

Two important insights regarding the role of model features and prior choice emerge. First, a long-run trend or t-distributed stochastic volatility, substantially improve forecast performance

Table 4: Monte Carlo Simulations: Dense Coefficient Vector

Prior	K	σ_k	$\overline{\sigma}$	RMSFE	RMSE($\mathcal{A}$)	RMSE($\mathcal{A}^C$)	MCC	TPR	FPR	RMSFE	CRPS
DGP 1: Steep MIDAS Coefficients											
HS	30	0.5	1.466	**0.056**	**0.066**	**0.015**	**0.097**	**0.549**	**0.447**	**2.645**	**1.883**
		0.9	1.613	**0.066**	**0.076**	**0.027**	**0.061**	**0.588**	**0.523**	**2.439**	**1.719**
	50	0.5	2.05	**0.068**	**0.08**	**0.021**	**0.086**	**0.499**	**0.407**	**4.068**	**3.244**
		0.9	2.004	**0.075**	**0.086**	**0.032**	**0.036**	**0.528**	**0.489**	**3.713**	**2.498**
	100	0.5	2.915	**0.073**	**0.085**	**0.022**	**0.051**	**0.443**	**0.389**	**6.514**	**6.157**
		0.9	2.76	**0.079**	**0.093**	**0.029**	**0.008**	**0.405**	**0.397**	**3.529**	**4.457**
GAL-SS	30	0.5	1.466	0.82	0.83	0.53	2.75	0.98	0.57	0.85	0.72
		0.9	1.613	1.04	1.03	0.94	2.06	0.56	0.40	0.84	0.64
	50	0.5	2.05	0.94	0.95	0.70	2.09	0.74	0.47	1.08	0.74
		0.9	2.004	0.99	0.99	0.96	2.06	0.45	0.35	0.95	0.74
	100	0.5	2.915	1.04	1.05	0.69	2.03	0.38	0.23	1.04	0.67
		0.9	2.76	1.07	1.07	0.81	9.76	0.36	0.22	0.91	0.41
GIGG	30	0.5	1.466	0.63	0.63	0.51	3.60	1.18	0.61	0.84	0.66
		0.9	1.613	0.80	0.82	0.57	2.50	0.85	0.64	0.84	0.64
	50	0.5	2.05	0.70	0.70	0.73	2.47	1.12	0.81	0.85	0.59
		0.9	2.004	0.81	0.82	0.71	2.62	0.84	0.70	0.91	0.67
	100	0.5	2.915	0.87	0.87	0.88	2.07	1.00	0.86	0.77	0.47
		0.9	2.76	0.91	0.92	0.81	5.93	0.96	0.85	0.92	0.42
DGP 3: Flat MIDAS Coefficients											
HS	30	0.5	0.824	**0.030**	**0.035**	**0.011**	**0.070**	**0.419**	**0.346**	**1.499**	**0.909**
		0.9	0.731	**0.036**	**0.040**	**0.018**	**0.049**	**0.443**	**0.394**	**1.460**	**0.975**
	50	0.5	1.034	**0.034**	**0.039**	**0.017**	**0.040**	**0.394**	**0.352**	**2.129**	**1.333**
		0.9	0.969	**0.045**	**0.049**	**0.03**	**0.011**	**0.454**	**0.442**	**1.360**	**1.037**
	100	0.5	1.434	**0.034**	**0.039**	**0.016**	**0.028**	**0.332**	**0.304**	**2.841**	**1.644**
		0.9	1.347	**0.041**	**0.046**	**0.025**	**0.010**	**0.354**	**0.343**	**3.370**	**1.876**
GAL-SS	30	0.5	0.824	0.93	0.93	0.86	2.80	1.22	0.88	0.79	0.80
		0.9	0.731	0.97	0.98	0.82	1.71	0.70	0.59	1.08	0.84
	50	0.5	1.034	0.97	0.98	0.87	2.87	0.98	0.76	0.82	0.70
		0.9	0.969	0.87	0.89	0.66	6.00	0.56	0.44	0.90	0.66
	100	0.5	1.434	1.10	1.09	1.14	2.19	0.70	0.59	0.79	0.76
		0.9	1.347	1.03	1.02	1.01	2.85	0.41	0.36	0.84	0.80
GIGG	30	0.5	0.824	0.84	0.84	0.78	3.34	1.25	0.79	0.77	0.77
		0.9	0.731	0.85	0.85	0.85	2.04	0.89	0.74	1.01	0.76
	50	0.5	1.034	0.89	0.90	0.80	2.67	1.08	0.89	0.83	0.73
		0.9	0.969	0.76	0.79	0.54	6.47	0.79	0.65	0.95	0.70
	100	0.5	1.434	0.99	1.00	0.89	2.18	1.01	0.90	0.77	0.74
		0.9	1.347	0.90	0.91	0.76	2.67	0.81	0.76	0.74	0.71

Notes: The table shows in bold the level of each of the evaluation metrics, while those of the GAL-SS and GIGG are in relative terms to the corresponding entry of the horseshoe model. $\overline{\sigma}$ is the Monte Carlo error of the DGP. These results refer to dense DGPs only.

relative to a simple BMIDAS. Second, the new shrinkage prior enhances nowcast performance by inducing group-wise sparsity while enabling the model to flexibly shift between signals. This feature proves particularly relevant once the Covid-19 pandemic is included into the analysis, since the heterogeneous nature of the shock calls for a nowcasting framework that can flexibly shift across signals from sparse roups of indicators, rather than relying on dense signals due to aggregate macroeconomic co-movement.

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.

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.

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

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

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.

Bitto, A. and S. Frühwirth-Schnatter (2019). Achieving shrinkage in a time-varying parameter model framework. *Journal of Econometrics 210*(1), 75–97.

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, 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.

Clements, M. P. and A. B. Galvão (2009). Forecasting us output growth using leading indicators: An appraisal using midas models. *Journal of Applied Econometrics 24*(7), 1187–1206.

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*.

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. 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.

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.

Jurado, K., S. C. Ludvigson, and S. Ng (2015). Measuring uncertainty. *American Economic Review 105*(3), 1177–1216.

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

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.

Zhang, Y. D., B. P. Naughton, H. D. Bondell, and B. J. Reich (2022). Bayesian regression using a prior on the model fit: The R2-D2 shrinkage prior. *Journal of the American Statistical Association 117*(538), 862–874.

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:

$$\begin{aligned}\overline{\theta}_{k,j} &= \Lambda_*((X'\Lambda_{t,h}^{-1}\boldsymbol{z}_{\boldsymbol{k}}^{(\boldsymbol{m})})^{-1} + \Lambda)\hat{\theta}_{k,j} \\ &= (1-\kappa_{k,j})\hat{\theta}_{k,j},\end{aligned} \tag{17}$$

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:

$$\begin{aligned}\pi(\kappa_k|\bullet) =& \frac{\Gamma(a_k + (p_k+1)b_k)}{\Gamma(a_k)\Gamma(b_k)^{k+1}} \prod_{j=1}^{p_k} \tilde{s}_{k,j}^{b_k} \left(1 + \sum_{j=1}^{p_k} \tilde{s}_{k,j}\frac{\kappa_{k,j}}{(1-\kappa_{k,j})}\right)^{-(a_k+(1+p_k)b_k)} \times \\ & \left(\prod_{j=1}^{p_k} \kappa_{k,j}^{b_k-1}(1-\kappa_{k,j})^{-(b_k+1)}\right),\end{aligned} \tag{18}$$

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

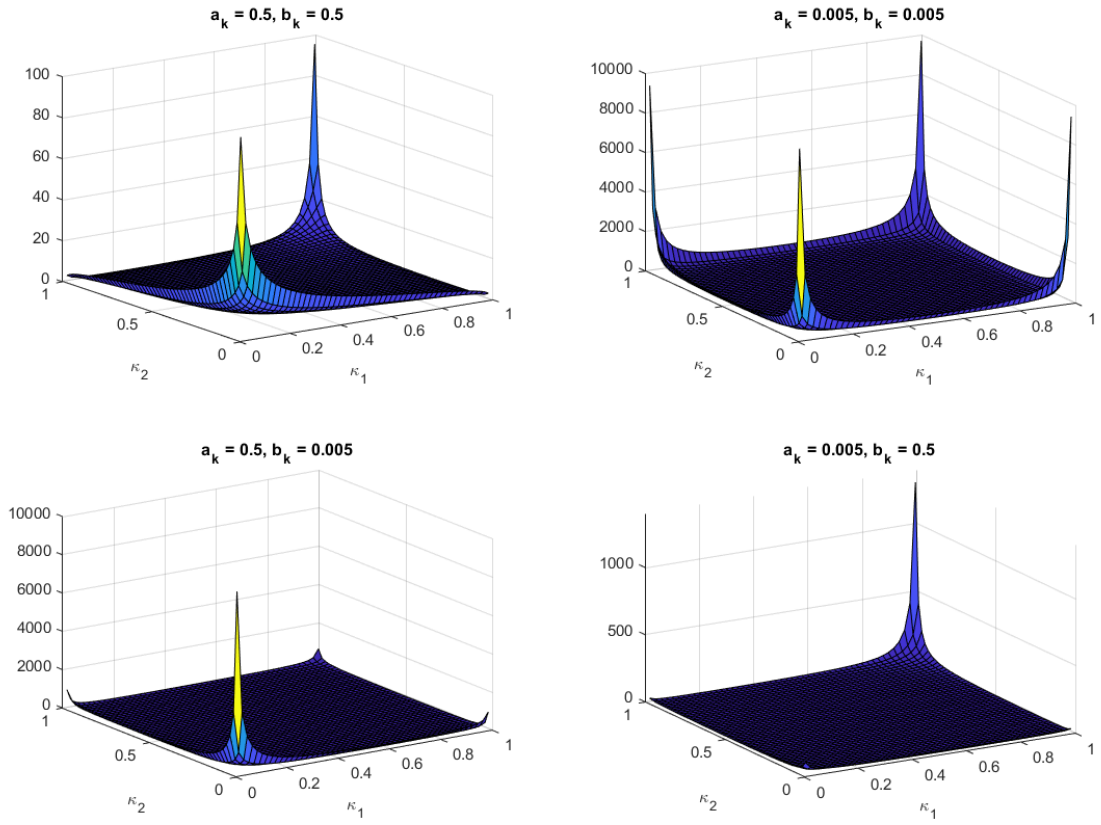


Figure B1: Bi-variate shrinkage coefficient plots for various hyper-parameter values. a_g controls group-level sparsitym while b_g controls the degree of correlation with the overall sparsity level.

(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:

$$(\vartheta|\boldsymbol{y}, \bullet) \sim IG(\frac{\sum_{k=1}^{K}(p_k + 1) + 1}{2}, \theta' \Lambda_p^{-1} \theta/2 + \frac{1}{\nu_p}) \tag{19}$$

$$(\gamma_k^{-2}|\boldsymbol{y},\bullet) \sim GIG(\frac{p_k+1}{2} - a_k, \frac{1}{\vartheta^2}\sum_{j=1}^{p_k+1}\frac{\theta_{kj}^2}{\varphi_{kj}^2}) \tag{20}$$

$$(\varphi_{kj}|\boldsymbol{y},\bullet) \sim IG(b_k + \frac{1}{2}, 1 + \frac{\theta_{kj}^2}{2\vartheta^2\gamma_k^2}) \tag{21}$$

$$(\nu_p|\boldsymbol{y},\bullet) \sim IG(1, \frac{1}{\vartheta}), \tag{22}$$

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:

$$\begin{aligned} &y_t = \tau_t + \theta' Z_{t-h}^{(m)} + \sqrt{\lambda_t} e^{\frac{1}{2}(h_0 + w_h \tilde{h}_t)} \tilde{\epsilon}_t^y, \\ &\tilde{\epsilon}_t^y \sim N(0,1),\ \lambda_t \sim IG(\nu/2, \nu/2) \end{aligned} \tag{23}$$

$$\begin{aligned} &\tau_t = \tau_{t-1} + e^{\frac{1}{2}(g_0 + w_g \tilde{g}_t)},\ \tilde{\epsilon}_t^g \sim N(0,1) \\ &\tilde{h}_t = \tilde{h}_{t-1} + \tilde{\epsilon}_t^h,\ \tilde{\epsilon}_t^h \sim N(0,1),\ \tilde{h}_0 = 0 \\ &\tilde{g}_t = \tilde{g}_{t-1} + \tilde{\epsilon}_t^g,\ \tilde{\epsilon}_t^g \sim N(0,1),\ \tilde{g}_0 = 0 \end{aligned} \tag{24}$$

$$\begin{aligned} &w_g \sim N(0, V_{w_g}),\quad w_h \sim N(0, V_{w_h}) \\ &h_0 \sim N(a_{0,h}, b_{0,h}),\quad g_0 \sim N(a_{0,g}, b_{0,g}) \\ &\tau_0 \sim N(a_{0,\tau}, b_{0,\tau}) \end{aligned} \tag{25}$$

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:

$$
\begin{aligned}
h_t &= h_0 + w_h \tilde{h}_t \\
g_t &= g_0 + w_g \tilde{g}_t
\end{aligned}
\tag{26}
$$

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 23 that then:

$$
\boldsymbol{y}|\boldsymbol{\tau}, \bullet \sim N(\boldsymbol{\tau}, \Lambda_h \Lambda_t^{1/2}). \tag{27}
$$

To derive the implicit prior on $\boldsymbol{\tau}$, start by vectorising the state process $\boldsymbol{\tau}$ in 24:

$$
H\tau = \tilde{\alpha}^{\boldsymbol{\tau}} + \eta^{\tau}, \tag{28}
$$

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 28, one can write the joint prior of $\boldsymbol{\tau}$ as:

$$
\boldsymbol{\tau}|\tau_0, \bullet \sim N(\tau 1_T, (H' \Lambda_g^{-1} H)^{-1}), \tag{29}
$$

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:

$$
\boldsymbol{\tau}|\boldsymbol{y}, \tau_0, \bullet \sim N(\hat{\boldsymbol{\tau}}, K_\tau^{-1}), \tag{30}
$$

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 25, the posterior is:

$$
\tau_0|\boldsymbol{y}, \bullet \sim N(\hat{\tau}_0, K_{\tau_0}^{-1}), \tag{31}
$$

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

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:

$$\boldsymbol{y}^+ = h_0 1_T + w_h \tilde{\boldsymbol{h}} + \tilde{\epsilon}^{y^+}. \tag{32}$$

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:

$$\tilde{\boldsymbol{h}}|\boldsymbol{y}, \bullet \sim N(\hat{\tilde{\boldsymbol{h}}}, K_{\tilde{h}}^{-1}), \tag{33}$$

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

The remaining conditional posteriors associated with $\boldsymbol{h}$ are those of h_0 and w_h. Notice from 32, 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:

$$\zeta_h|\boldsymbol{y}, \bullet \sim N(\hat{\zeta}_h, K_{\zeta,h}^{-1}), \tag{34}$$

$K_{\zeta_h} = (b_{0,\zeta_h}^{-1} + X_h' \Omega_s^{-1} X_h)$ 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:

$$p(\boldsymbol{\lambda}|\bullet) \propto \prod_{t=1}^{T} \lambda_t^{-\frac{\nu+1}{2}+1} e^{-\frac{1}{2\lambda_t}(\nu + \frac{(y_t - \tau_t - \theta' Z_{t-h}^{(m)})^2}{exp(h_t)})}. \tag{35}$$

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

$$\lambda_t \sim \mathcal{G}^{-1}(\frac{\nu+1}{2}, \frac{1}{2}\frac{(y_t - \tau_t - \theta' Z_{t-h}^{(m)})^2}{exp(h_t)}) \tag{36}$$

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:

$$\begin{aligned} p(\nu|\bullet) &\propto p(\boldsymbol{\lambda} p(\nu)) \\ &\propto \prod_{t=1}^{T} \frac{(\nu/2)^{\frac{\nu}{2}}}{\Gamma(\nu/2)} \lambda_t^{-(\frac{\nu}{2}+1)} e^{-\frac{\nu}{2\lambda_t}} \\ &= \frac{(\nu/2)^{\frac{T}{2}}}{\Gamma(\nu/2)^T} (\prod_{t=1}^{T})^{-(\frac{\nu}{2}+1)} e^{\frac{\nu}{2}\sum_{t=1}^{T}\lambda_t^{-1}} \end{aligned} \tag{37}$$

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

$$\alpha(X, Y) = min\left\{ \frac{f(Y)g(X)}{f(X)g(Y)}, 1 \right\}, \tag{38}$$

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 37

We iterate sampling steps 1.-7. initially for 5000 times for burnin 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:

$$\mathcal{L}(\tilde{\boldsymbol{Y}}, \alpha) = \frac{1}{2}||\boldsymbol{Z}\alpha - \tilde{\boldsymbol{Y}}||_2^2 + \sum_{k=1}^K \phi_k ||\alpha_k||_2, \tag{39}$$

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]:

$$\begin{aligned} \mathcal{L}(\tilde{\boldsymbol{Y}}, \alpha) &= E_{\tilde{\boldsymbol{Y}}|\bullet}[\tilde{\boldsymbol{Y}}, \alpha|\bullet] \\ &= \sum_{k=1}^K \phi_k ||\alpha_k||_2 + \frac{1}{2}||\boldsymbol{Z}\alpha - \boldsymbol{Z}\theta||_2^2 + \frac{1}{2}tr(\Sigma) \end{aligned} \tag{40}$$

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

$$\begin{aligned}\mathcal{L}(\theta,\alpha) &= E_{\theta|\boldsymbol{y}}[\mathcal{L}(\tilde{\boldsymbol{Y}},\alpha)] \\ &= \frac{1}{2}||\boldsymbol{Z}\alpha - \boldsymbol{Z}\overline{\theta}||_2^2 + \frac{1}{2}tr(\boldsymbol{Z}'\boldsymbol{Z}\Sigma_\theta) + \sum_{k=1}^{K}\phi_k||\alpha_k||_2,\end{aligned} \tag{41}$$

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

$$\mathcal{L}(\theta,\alpha) = \frac{1}{2}||\boldsymbol{Z}\alpha - \boldsymbol{Z}\overline{\theta}|| + \sum_{k=1}^{K}\phi_k||\alpha_k||_2. \tag{42}$$

Notice, that we follow Chakraborty et al. (2020); Huber et al. (2019) by solving 41 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:

$$\alpha_k^l = (||r_k^{(l-1)}||_2^2 - \phi_k)_+ \frac{r_k^{(l-1)}}{||r_k^{(l-1)}||_2^2},\ l = 1,\cdots,L, \tag{43}$$

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

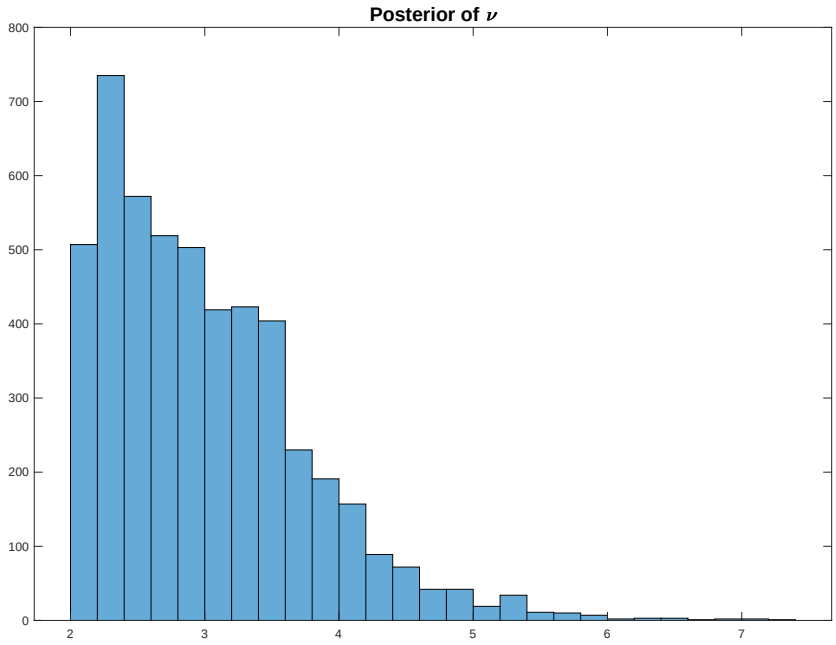


Figure B2: Posterior Degrees of Freedom of the t-distribution, based on the entire in-sample and last nowcast period.

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

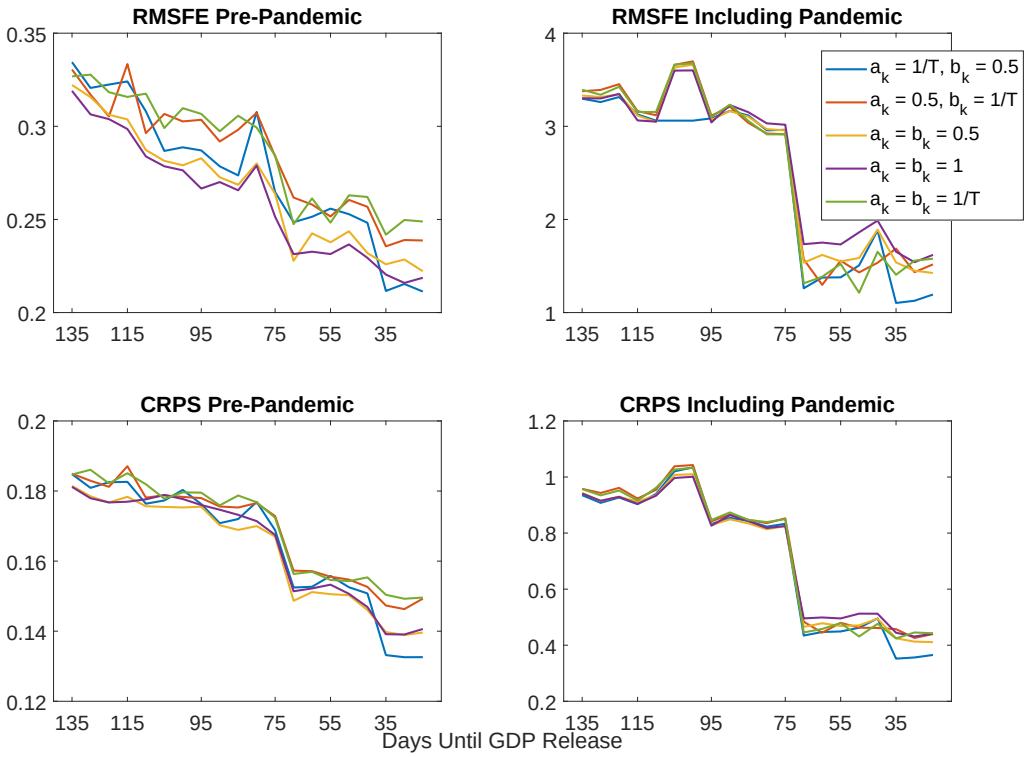


Figure B3: RSMFE and CRPS pre-pandemic and including pandemic for the Trend-SV-t-GIGG model with different hyper-parameter combinations. $1/T$ is adjusted to the in-sample length at each quarter the nowcasts are conducted. The GIGG prior implemented in the empirical application has $a_k = 1/T, b_k =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

B.4 Inclusion Probabilities

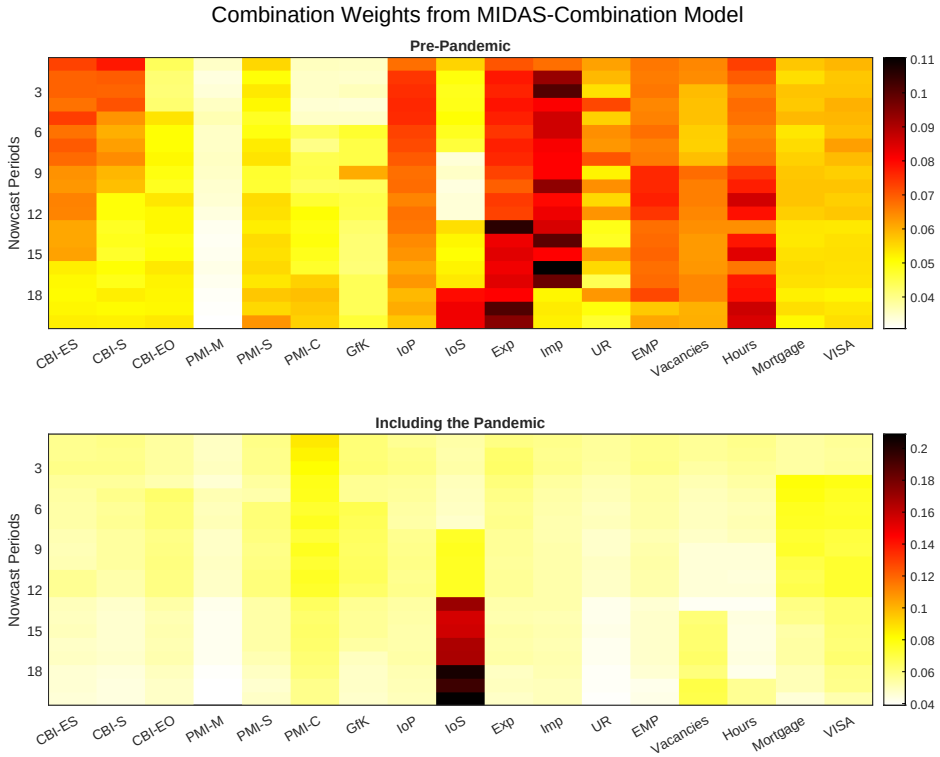


Figure B4: Combination weights of the Combination model of high-frequency variables. Weights pre-pandemic don't change much quarter to quarter, so are averages across quarters until Q4-2019. Each subsequent panel shows the weights for each quarter separately. Vertical axis in each panel represents a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