

NOWCASTING GROWTH USING GOOGLE TRENDS DATA: A BAYESIAN STRUCTURAL TIME SERIES MODEL

Arnab Bhattacharjee

Edinburgh Business School, Heriot-Watt University and NIESR

David Kohns

Heriot-Watt University

About the National Institute of Economic and Social Research

The National Institute of Economic and Social Research is Britain's longest established independent research institute, founded in 1938. The vision of our founders was to carry out research to improve understanding of the economic and social forces that affect people's lives, and the ways in which policy can bring about change. Over eighty years later, this remains central to NIESR's ethos. We continue to apply our expertise in both quantitative and qualitative methods and our understanding of economic and social issues to current debates and to influence policy. The Institute is independent of all party-political interests.

National Institute of Economic and Social Research

2 Dean Trench St

London SW1P 3HE

T: +44 (0)20 7222 7665

E: enquiries@niesr.ac.uk

www.niesr.ac.uk

Registered charity no. 306083

This paper was first published in May 2022

Competing Interest Statement: The views expressed in this paper are based on research and are not attributed to the organizations to which the researchers are affiliated. There are no conflicts of interest. The usual disclaimer applies.

© National Institute of Economic and Social Research 2022

Nowcasting Growth using Google Trends Data: A Bayesian Structural Time Series Model

Arnab Bhattacharjee and David Kohns

Abstract

This paper investigates the benefits of internet search data in the form of Google Trends for nowcasting real U.S. GDP growth in real time through the lens of mixed frequency Bayesian Structural Time Series (BSTS) models. We augment and enhance both model and methodology to make these better amenable to nowcasting with large number of potential covariates. Specifically, we allow shrinking state variances towards zero to avoid overfitting, extend the SSVS (spike and slab variable selection) prior to the more flexible normal-inverse-gamma prior which stays agnostic about the underlying model size, as well as adapt the horseshoe prior to the BSTS. The application to nowcasting GDP growth as well as a simulation study demonstrate that the horseshoe prior BSTS improves markedly upon the SSVS and the original BSTS model with the largest gains in dense data-generating-processes. Our application also shows that a large dimensional set of search terms is able to improve nowcasts early in a specific quarter before other macroeconomic data become available. Search terms with high inclusion probability have good economic interpretation, reflecting leading signals of economic anxiety and wealth effects.

Classification: C11, C22, C55, E37, E66

Keywords: Global-Local Priors, Non-Centred State Space, Shrinkage, Google Trends

Nowcasting Growth using Google Trends Data: A Bayesian Structural Time Series Model*

David Kohns[†]

Economics, Heriot-Watt University, UK

and

Arnab Bhattacharjee

Economics, Heriot-Watt University, UK &

National Institute of Economic & Social Research, UK.

May 13, 2022

Abstract

This paper investigates the benefits of internet search data in the form of Google Trends for nowcasting real U.S. GDP growth in real time through the lens of mixed frequency Bayesian Structural Time Series (BSTS) models. We augment and enhance both model and methodology to make these better amenable to nowcasting with large number of potential covariates. Specifically, we allow shrinking state variances towards zero to avoid overfitting, extend the SSVS (spike and slab variable selection) prior to the more flexible normal-inverse-gamma prior which stays agnostic about the underlying model size, as well as adapt the horseshoe prior to the BSTS. The application to nowcasting GDP growth as well as a simulation study demonstrate that the horseshoe prior BSTS improves markedly upon the SSVS and the original BSTS model with the largest gains in dense data-generating-processes. Our application also shows that a large dimensional set of search terms is able to improve nowcasts early in a specific quarter before other macroeconomic data become available. Search terms with high inclusion probability have good economic interpretation, reflecting leading signals of economic anxiety and wealth effects.

Keywords: Global-Local Priors, Non-Centred State Space, Shrinkage, Google Trends

*The authors thank Atanas Christev, Marco Del Negro, Sylvia Frühwirth-Schnatter, George Kapetanios, Sylvia Kaufmann, Gary Koop, Stephen Millard, Aubrey Poon, Galina Potjagailo, Mark Schaffer, Herman van Dijk, and all participants of the Annual Workshop on Financial Econometrics in Orebo (2019), the Bank of England Big Data Conference (2019), the CEF conference (2020), the International Symposium on Forecasting (2019,2020), the Panmure House PhD conference (2019), the Study Center Gerzensee and Norges Bank conferences for their invaluable feedback. The constructive suggestions and comments from two anonymous reviewers and encouragement from the Editor helped us revise and improve upon the paper substantially. Their help is gratefully acknowledged. The usual disclaimer applies.

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

1 Introduction

The primary object of nowcasting models is to produce ‘early’ forecasts of target variables associated with long delays in data publication by exploiting the real time data publication schedule of the explanatory data set. While prediction is the primary goal here, the selected models can also sometimes provide structural interpretations *ex post*. Nowcasting is particularly relevant to central banks and other policymakers who are tasked with conducting forward looking policies on the basis of key economic variables such as GDP or inflation. Inflation data are, however, published with a lag of up to 7 weeks with respect to their reference period, and precise estimates of GDP can take years.¹ Since even monthly macroeconomic data arrive with considerable lag, it is now common to combine, next to traditional macroeconomic data, ever more information from Big Data sources such as internet search terms, satellite data, scanner data, etc. which have the advantage of being available in near real time (Bok et al., 2018). The recent Covid-19 pandemic has given further impetus to this trend, as faster indicators have proven especially useful in modelling the unprecedentedly sharp movements in the economy that traditional macroeconomic models fail to capture in a timely manner (Antolin-Diaz et al., 2021; Woloszko, 2020).

In this paper we add to the burgeoning literature on using Google search data in the form of Google Trends (GT), which measure the relative search volume of certain search terms entered into the Google search engine, to nowcast aggregate economic time-series. In particular, we investigate the benefits of using monthly Google search information for nowcasting quarterly U.S. real GDP growth in real-time compared to traditional macro data and survey information. We contribute to this literature by being, to our best knowledge, the first paper to investigate the benefit of search information above and beyond macroeconomic data for the U.S. including performance during the Covid-19 pandemic period.

To deal with the specificities of the data, we propose robust nowcasting methods that are amenable to situations in which the policymaker needs to combine traditional and non-traditional data sources while providing tractable variable selection properties. For this purpose, we adapt current generation state space and regression priors to the widely popular Bayesian Structural Time Series model (BSTS) (Scott and Varian, 2014). Results from our nowcasting

¹The exact lag in publications of GDP and inflation depends as well on which vintage of data the econometrician wishes to forecast. Since early vintages of aggregate quantities such as GDP can display substantial variation between vintages, this is not a trivial issue.

application show that Google’s search information improves nowcasts of GDP growth, particularly early on in the quarter before macroeconomic data are published. We show that our extensions allow for accuracy gains of up to 40% during certain nowcast periods in point as well as in density nowcasts compared to the original BSTS model of Scott and Varian (2014) while retaining its interpretability. These results are confirmed in a simulation study which checks robustness to a variety of data-generating processes.

In the following, we firstly discuss the state and regression priors as well as posteriors for our extended BSTS models and provide efficient sampling algorithms. In section 3, we elaborate further on the data used for nowcasting, including dealing with mixed frequency, the data publication calendar and the specificities of the Google Trends data set. In section 4, we present results based on our empirical application of nowcasting U.S. GDP growth, which is followed in section 5 by the results from our simulation study. Finally, Section 6 concludes with a discussion and avenues for future research.

2 Bayesian Structural Time Series

2.1 The Original Model

The Bayesian Structural Time Series (BSTS) model, as proposed by Scott and Varian (2014), provides a conceptually attractive model for nowcasting aggregate economic time-series with heterogeneous data sources, as it flexibly estimates latent time-trends, seasonality and deviations or ‘irregular’ dynamics through variable selection using a high-dimensional shrinkage prior. Denote the target variable to be nowcast by $y_t = (y_1, \dots, y_T)'$ and the K-dimensional explanatory data set as $x_t = (x'_1, \dots, x'_T)'$ which for now are sampled at the common frequency, t . Then our model is as follows:

$$\begin{aligned}
y_t &= \tau_t + x'_t \beta + \delta_t + \epsilon_t, \epsilon_t \sim N(0, \sigma_y^2) \\
\tau_t &= \tau_{t-1} + \alpha_t + \epsilon_t^\tau, \epsilon_t^\tau \sim N(0, \sigma_\tau^2) \\
\alpha_t &= \alpha_{t-1} + \epsilon_t^\alpha, \epsilon_t^\alpha \sim N(0, \sigma_\alpha^2) \\
\delta_t &= - \sum_{s=1}^{S-1} \delta_{t-s} + \epsilon_t^\delta, \epsilon_t^\delta \sim N(0, \sigma_\delta^2).
\end{aligned} \tag{1}$$

(1) is a linear state space model with Gaussian errors and states $\{\tau_t, \alpha_t, \delta_t\}_{t=1}^T$ which capture long-run trends and S seasonal components δ_t . The deviation from τ_t describes variation from

a long-run trend which, when applied to the level of GDP can be interpreted as the output gap (Watson, 1986; Grant and Chan, 2017). α_t allows for a drift term in the trend which is often observed in stock variables such as in GDP, aggregate consumption and inflation (Grant and Chan, 2017; Chan, 2017b).

Variable selection on the possibly high-dimensional $K \times 1$ response vector β in the BSTS model of Scott and Varian (2014) is done via a two component conjugate spike-and-slab prior. Estimation is standard (George and McCulloch, 1993), and states τ_t , α_t and δ_t are estimated jointly via the forward filtering backward sampling (FFBS) algorithm of Durbin and Koopman (2002) based on the Kalman filter. This implementation relies on Normal-Inverse Gamma (N-IG) priors for the states and state variances for conditional conjugacy. While the BSTS model is a natural model for many time-series applications, we bring 3 important methodological innovations which make it more robust to overfitting trend estimation and variable selection with heterogeneous high dimensional data.

2.2 Model enhancements

2.2.1 Non-Centred Bayesian Structural Time Series

In line with previous nowcasting studies, this paper focuses on nowcasting GDP *growth* rather than levels. However, two problems arise when applying model (1) directly to growth variables. As growth variables are often approximately stationary, conceptually, the inclusion of α_t implies that GDP growth follows a boundless drift for which there is little structural justification or empirical evidence. A non-drifting stochastic trend, on the other hand, has been shown to markedly improve nowcasts of GDP growth as shown in Antolin-Diaz et al. (2017), especially when the state variances are tightly controlled by priors such that the stochastic trend does not wander too wildly. This suggests that modeling time-variation is preferred over de-trending a priori.² The underlying rationale for this improvement is the well known empirical finding of changes in long-run GDP growth (Kim and Nelson, 1999; McConnell and Perez-Quiros, 2000; Jurado et al., 2015). Econometrically, the additional problem is that the IG priors with no prior mass on zero, as implemented for Bayesian linear state space methods, can bias posterior state variances away from zero, thereby potentially leading to false support for state dynamics which can hurt forecast performance.

²Modelling an I(1) component in U.S. GDP growth is additionally consistent with Harvey's local-linear trend model (Harvey, 1985), the Hodrick and Prescott (1997) filter and Stock and Watson (2012).

We extend model (1) to flexibly let the data shut down state dynamics, and therefore broaden the applicability of model (1), by adopting the non-centred parameterisation of the state space as suggested by Frühwirth-Schnatter and Wagner (2010). The non-centred parameterisation models state variances directly in the observation equation, which with normal priors, exerts much stronger shrinkage than IG priors.³ This allows additionally for valid inference on testing for zero posterior variance via Savage-Dickey density ratios, as will be further discussed in section 5. Testing for zero posterior variance would be very challenging in a frequentist hypothesis testing approach because the null hypothesis of constant state in model (1) lies on the boundary of the parameter space.

The non-centred model considered for the empirical application is equivalently written as:

$$y_t = \tau_0 + \sigma_\tau \tilde{\tau}_t + t\alpha_0 + \sigma_\alpha \sum_{s=1}^t \tilde{\alpha}_s + x'_t \beta + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma^2) \quad (2)$$

and

$$\begin{aligned} \tilde{\tau}_t &= \tilde{\tau}_{t-1} + \tilde{u}_t^\tau, \quad \tilde{u}_t^\tau \sim N(0, 1) \\ \tilde{\alpha}_t &= \tilde{\alpha}_{t-1} + \tilde{u}_t^\alpha, \quad \tilde{u}_t^\alpha \sim N(0, 1) \end{aligned} \quad (3)$$

with starting values $\tilde{\tau}_0 = \tilde{\alpha}_0 = 0$. Note that the seasonal component is left out for estimation due to the small sample length of the Google Trends data set and differing seasonal patterns between monthly and quarterly data.⁴ To see that (2) and (3) is equivalent to (1), let:

$$\begin{aligned} \alpha_t &= \alpha_0 + \sigma_\alpha \tilde{\alpha}_t \\ \tau_t &= \tau_0 + \sigma_\tau \tilde{\tau}_t + t\alpha_0 + \sigma_\alpha \sum_{s=1}^t \alpha_s \end{aligned} \quad (4)$$

Hence, by setting $y_t = \tau_t + x'_t \beta + \epsilon_t$, it is clear that

$$\begin{aligned} \alpha_t - \alpha_{t-1} &= \sigma_\alpha (\tilde{\alpha}_t - \tilde{\alpha}_{t-1}) \\ &= \sigma_\alpha + \tilde{u}_t^\alpha \\ \tau_t - \tau_{t-1} &= \alpha_0 + \sigma_\alpha \tilde{\alpha}_t + \sigma_\tau (\tilde{\tau}_t - \tilde{\tau}_{t-1}) \\ &= \alpha + \sigma_\tau + \tilde{u}_t^\tau \end{aligned} \quad (5)$$

which recovers (1). Since $\sigma_{\tau, \alpha}$ are allowed to have support on the real line, they are not identified in multiplication with the states: the likelihood is invariant to signs of σ_α and σ_τ . Consequently,

³Formally, a normal prior on the state standard deviation can be shown to imply a Gamma prior on the state variance.

⁴For further discussion, please see section 3

mixing of the posterior state standard deviations can be poor and their distributions are likely to be bi-modal (Frühwirth-Schnatter and Wagner, 2010). This issue is addressed by randomly permuting signs in the Gibbs sampler as explained below. Similar to Frühwirth-Schnatter and Wagner (2010), we assume normal priors centred at 0 for $\sigma_i : \sigma_i \sim N(0, V_i) \forall i \in \{\tau, \alpha\}$.

Collecting all state space parameters in $\theta = (\tau_0, \alpha_0, \sigma_\tau, \sigma_\alpha)$, we assume an independent multivariate normal prior with diagonal covariance matrix:

$$\theta \sim N(\theta_0, V_\theta). \quad (6)$$

While the state processes $\{\tilde{\tau}, \tilde{\alpha}\}_{t=1}^T$ can be estimated by any state space algorithm, we opt for the precision sampler method of Chan and Jeliazkov (2009) which is outlined in Appendix (A.4.1) along with the state posteriors. In contrast to FFBS type algorithms, it samples the states without recursive estimation which speeds up computation significantly.

2.2.2 SSVS Prior

Our second enhancement concerns the SSVS prior. Variable selection in the BSTS model of Scott and Varian (2014) is done via a two component conjugate spike-and-slab prior which utilises a variant of Zellner's g-prior and fixed expected model size. While computationally fast due to conjugacy, many high-dimensional problems benefit from prior independence (Moran et al., 2018) and a fully hierarchical formulation to let the data decide on the most likely value of the parameters (Ishwaran et al., 2005).

Therefore, we follow Ishwaran et al. (2005)'s extension to the SSVS prior, the Normal-Inverse-Gamma prior:

$$\begin{aligned} \beta_j | \gamma_j, \delta_j^2 &\sim \gamma_j N(0, \delta_j^2) + (1 - \gamma_j) N(0, c \times \delta_j^2) \\ \delta_j^2 &\sim \text{Gamma}(a_1, a_2) \\ \gamma_j &\sim \text{Bernoulli}(\pi_0) \\ \pi_0 &\sim \text{Beta}(b_1, b_2) \end{aligned} \quad (7)$$

where $j \in (1, \dots, K)$. The intuition remains the same as compared to the spike-and-slab prior of Scott and Varian (2014) in that the covariate's effect is modeled by a mixture of normals where it is either shrunk close to zero via a narrow distribution around zero (the spike component) or estimated freely though a relatively diffuse normal distribution (the slab component). Sorting into each component is handled through an indicator variable, γ_j , and the hyperparameter c is chosen to be a very small number, thereby forcing shrinkage of noise variables to close

to zero. While in the original BSTS model, the indicator variable, γ_j , depends on a fixed prior π_0 which governs the prior inclusion probability of a variable, (7) allows for it to be estimated from the data through another level of hierarchy. We set $b_1 = b_2 = 1$, which effectively assumes that any expected model size is a priori possible and thus allows for sparse but also dense model solutions as recommended by Giannone et al. (2021). Finally, the prior variance δ_j^2 is also allowed to be hierarchical. Posteriors are standard and described in the Appendix (A.1). The posterior of γ_j is of special interest to the analyst as it gives a data informed measure of importance of a variable. Specifically, $p(\gamma|y)$ can be interpreted as the posterior inclusion probability of a variable.

2.2.3 Horseshoe Prior

Our third and final enhancement of the BSTS models extends the employed shrinkage priors to the horseshoe prior. Like many recently popularised shrinkage priors, the horseshoe prior belongs to the broader class of global-local priors which take the following general form:

$$\begin{aligned} \beta_j | \lambda_j^2, \nu^2, \sigma^2 &\sim N(0, \lambda_j^2 \nu^2 \sigma^2), j \in (1, \dots, K) \\ \lambda_j &\sim \pi(\lambda_j) d\lambda_j, j \in (1, \dots, K) \\ \nu &\sim \pi(\nu) d\nu \\ \sigma^2 &\sim \pi(\sigma^2) d\sigma^2 \end{aligned} \tag{8}$$

The idea of this family of priors is that the global scale ν controls the overall shrinkage applied to the regression, while the local scale λ_j allows for the local possibility of regressors to escape shrinkage if they have large effects on the response. A variety of global-local shrinkage priors have been proposed (Polson and Scott, 2010), but here we focus on arguably the most popular, the horseshoe prior of Carvalho et al. (2010) which employs two half Cauchy distributions for ν and λ_j :

$$\begin{aligned} \lambda_j &\sim C_+(0, 1) \\ \nu &\sim C_+(0, 1) \end{aligned} \tag{9}$$

These two fat tailed scale distributions imply a shrinkage profile that has the spike-and-slab prior in its limit and therefore offers a continuous approximation to the SSVS (Piironen et al., 2017) (see section A.2 in the appendix for further discussion). An additional attractive feature of the horseshoe prior is that it is completely automatic with respect to its hyperparameters and

has been shown to be excellent at forecasting in several previous studies (Huber et al., 2019, 2020; Cross et al., 2020; Follett and Yu, 2019). Due to its special connection to frequentist shrinkage priors (Polson and Scott, 2010), it offers not only good finite sample performance but also favourable asymptotic behaviour compared to competing global priors (Bhadra et al., 2019). Chakraborty et al. (2020) in particular show that the fractional posterior mean as a point estimator is rate optimal in the minimax sense using (9).

Nevertheless, fitting the horseshoe prior can be challenging when the scale parameters are not strongly identified by the data, which is particularly critical in cases where the likelihood is flat, for example, separable data in logistic regression (Piironen et al., 2017).⁵ We provide in the appendix a (A.2) robustness check based on Piironen et al. (2017) that are able to alleviate any identifiability concerns for the empirical study below.

Posteriors are described in the appendix (A.1).

2.2.4 SAVS Algorithm

Although the horseshoe prior shrinks noise variables towards zero, the importance of a variable for nowcasts may not be immediately clear from posterior summary statistics of the coefficients, especially when the posterior is multi-modal. To aid interpretability and simultaneously preserve predictive ability, we employ the signal adaptive variable selection (SAVS) algorithm of Ray and Bhattacharya (2018) to the posterior coefficients on a draw-by-draw basis. The algorithm uses a useful heuristic, inspired by frequentist lasso estimation, to threshold posterior regression coefficients to zero:

$$\phi_j = \text{sign}(\hat{\beta}_j) ||X_j||^{-2} \max(|\hat{\beta}_j| ||X_j|| - \kappa_j, 0), \quad (10)$$

where $X_j = (x_{j1}, \dots, x_{jT})'$ is the j^{th} column of the regressor matrix X, $\text{sign}(x)$ returns the sign of x and $\hat{\beta}$ represents a draw from the regression posterior. The parameter κ_j in (10) acts as a threshold for each coefficient akin to the penalty parameter in lasso regression which can be selected via cross-validation. Ray and Bhattacharya (2018) propose

$$\kappa_j = \frac{1}{|\hat{\beta}_j|^2}, \quad (11)$$

which ranks the coefficients inverse-squared proportionally and provides good performance compared to alternate penalty levels (Ray and Bhattacharya, 2018; Huber et al., 2019). To see the

⁵We thank an anonymous reviewer for having facilitated this discussion.

similarity to lasso style regularisation, the solution to (10) can be obtained by the following minimisation problem which is closely related to the adaptive lasso (Zou, 2006):

$$\bar{\phi} = \operatorname{argmin}_{\phi} \left\{ \frac{1}{2} \|X\hat{\beta} - X\phi\| + \sum_{j=1}^K \kappa_j |\phi_j| \right\}. \quad (12)$$

Here, $\bar{\phi}$ is the sparsified regression vector. Analogous to the SSVS posterior, the relative frequency of non-zero entries in the posterior coefficient vector can be interpreted as posterior inclusion probabilities. Integrating over the uncertainty of the parameters, we obtain the predictive distribution $p(\tilde{y}|y)$, which is similar to a Bayesian Model Averaged (BMA) posterior (Huber et al., 2019).

2.2.5 Sampling Algorithm

With the conditional posteriors for the regression and state components at hand (see Appendix A.1), we sample states as well as regression parameters with the following Gibbs sampler:

1. Sample $(\tilde{\tau}, \tilde{\alpha}|y, \theta, \beta, \sigma_y^2)$
2. Sample $(\theta|y, \beta, \tilde{\tau}, \tilde{\alpha}, \sigma_y^2)$
3. Randomly permute signs of $(\tilde{\tau}, \tilde{\alpha})$ and $(\sigma_\tau, \sigma_\alpha)$
4. Sample $(\beta|y, \theta, \tilde{\tau}, \tilde{\alpha}, \sigma_y^2)$
5. Sample $(\sigma_y^2|y, \tilde{\tau}, \tilde{\alpha}, \sigma_y^2)$

As mentioned in Section 2.2.1, states are sampled in a non-recursive fashion which exploits sparse matrix computation and precision sampling. The exact sampling algorithm is given in Appendix A.4.1. As discussed in Section 2.2.1, after sampling θ in step 2, we randomly permute signs of $(\tilde{\tau}, \tilde{\alpha}), (\sigma_\tau, \sigma_\alpha)$ to aid mixing. Step 4 of the sampler will depend on the prior and its respective hyperpriors. While the posterior sampling scheme for the SSVS is standard, we use the efficient posterior sampler of Bhattacharya et al. (2016) to sample the regression coefficients of the horseshoe prior. Compared to Cholesky based sampling as used for the SSVS, computation speed is markedly improved; see Appendix A.1.1. Note that in step 4, we perform SAVS sparsification via (10) on an iteration basis.

3 Data

3.1 Mixing Frequencies

In this paper, we relate monthly macro data commonly used for nowcasting based on Giannone et al. (2016) and internet search information via U-MIDAS skip-sampling to real quarterly U.S. GDP growth. The U-MIDAS approach to mixed frequency belongs to the broader class of ‘partial system’ models (Ba  bura et al., 2013), which directly relate higher frequency information to the lower frequency target variable by vertically realigning the covariate vector. The benefit of this mixed frequency method compared to restricted MIDAS and full system state space methods is its simplicity in that existing models and priors can directly be applied to U-MIDAS sampled data as well as its competitive performance, especially when the frequency mismatch between the target and the regressors is small (Foroni et al., 2015; Foroni and Marcellino, 2014), as is the case in our application. Switching notation from equation (1) to make it explicit that y_t is quarterly while x_t is sampled at a higher, i.e., monthly frequency, denote $x_{t,M} = (x_{1,t,M}, \dots, x_{K,t,M})$ and $\beta_m = (\beta_{1,M}, \dots, \beta_{K,M})'$ where $M = (1, 2, 3)$ denotes the monthly observation of the covariate within quarter, t . By concatenating each monthly column, we obtain a $T \times 3K$ regressor matrix \mathbf{X} and a $3K \times 1$ regression coefficient vector $\boldsymbol{\beta}$. This vertical realignment is visualised for a single representative regressor below:

$$\left(\begin{array}{c|ccc} y_{1stquarter} & x_{Mar} & x_{Feb} & x_{Jan} \\ y_{2ndquarter} & x_{Jun} & x_{May} & x_{Apr} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \end{array} \right) \quad (13)$$

3.2 Macroeconomic Data

The macro data set pertains to an updated version of the database of Giannone et al. (2016) (henceforth, ‘macro data’) which contains 13 time series which are closely watched by professional and institutional forecasters including real indicators (industrial production, house starts, total construction expenditure etc.), price data (CPI, PPI, PCE inflation), financial market data (BAA-AAA spread) and credit, labour and economic uncertainty measures (volume of commercial loans, civilian unemployment, economic uncertainty index etc.). We augment this data set with the composite Purchasing Managers Index (PMI) and the University of Michigan

Consumer Confidence Index (UMCI). These are often used as leading indicators for producer and consumer sentiment, respectively. The target variable for this application is deseasonalised U.S. real GDP growth (GDP growth) data as reported in the FRED dataset.^{6,7}

As early data vintages of macroeconomic data and GDP figures can exhibit substantial variation compared to final vintages (Croushore, 2006; Sims, 2002), there is no unambiguous choice of variable in evaluating nowcast models on historical data. Further complications can arise through changing definitions or methods of measurements (Carriero et al., 2015). In order to judge the expected performance of the proposed models from a real-time perspective, we only make use of the latest vintages of the series available at the point in time of the nowcast.

⁸ The stylised release calendar (1) simulates the data availability within the months during which nowcasts are conducted. For instance, at the 24th nowcast period, all data which became available during periods 1-24 will be updated according to their latest available vintages dating prior to the release of PCE and PCEPI, which are published typically during the last week of a given month. Real-time vintages are downloaded from the FRED database using the ‘FredFetch’ Matlab package.⁹

3.3 Google Trends

Google Trends (GT) are indices produced by Google on the relative search volume popularity of a given search term, search topic or pre-specified search category, conditional on a given time frame and location. The difference between individual search terms and topics/categories is that the latter measures the search popularity for a basket of search terms which are content-wise related to the specified topic or category. In particular, categories are further split into a 5-level hierarchy of categories which are fixed a priori,¹⁰ and topics can be assembled depending on the

⁶Here, the deseasonalisation pertains to the X13-ARIMA method and was performed prior to download from the FRED-MD website.

⁷We thank an anonymous reviewer who brought to our attention that instead of mixing pre-deseasonalisation techniques between macroeconomic data and Google Trends discussed below, one could also deseasonalise with common techniques such as the Loess filter. In doing so, the results in this paper remain qualitatively identical. Details are available upon request.

⁸The only exception is real GDP growth, for which, following previous nowcast studies (Carriero et al., 2015; Clark, 2011), we use the second vintage for nowcast evalutation.

⁹The Matlab package is available from <https://github.com/MattCocci/FredFetch>. PMI data were downloaded from [quandl.com](https://www.quandl.com) using Quandl code ‘ISM/MANPMI’.

¹⁰For an overview of categories and sub-categories, please see <https://github.com/pat310/google-trends-api/wiki/Google-Trends-Categories>

term one is interested in. For example, the user can specify the topic ‘Recession’, whose related search queries contain, among others, ‘recession’, ‘downturn’, and ‘economic depression’. Likewise, the category ‘Welfare & Unemployment’ relates to search queries about ‘unemployment’, ‘food stamps’ and ‘social security office’. A large literature on using individual Google Trends search terms¹¹ have shown that these data can improve predictions for economic time-series which have a clear connection to the specific search term used, such as using ‘unemployment benefits’ to predict unemployment (Smith, 2016). However, this approach has two potential limitations.

First, using broad search terms to capture general macroeconomic activity bears the risk of capturing spurious search behaviour. For example, the search term ‘jobs’ might contain search volume for ‘Steve Jobs’. Second, since many search terms will be related to multiple topics, there may be lack of interpretability. To reduce search term ambiguity and interpretability in relationship to real GDP growth, we use Google topics and categories instead of individual search terms, and choose these based on their relationship with various aspects of the economy. As forcefully argued by Woloszko (2020) and Fetzer et al. (2020), these mostly alleviate spuriously correlated search terms as the user can confine the search purpose. This is benefited by the fact that Google refines this basket of search terms, by taking into account where users click after the search has been conducted Woloszko (2020). Further, categories and topics can be conceptualised as factors based on search terms with the same meaning/purpose. Although the exact basket of search terms corresponding to a topic/category is not a priori accessible to the user, any topic or category with little predictive power will ultimately be shrunk to zero via the shrinkage priors employed in the proposed models.

Our sample comprises 37 Google Trends which were chosen based on capturing activity in various parts of the economy ranging from crisis/recession, labour market, personal finance, consumption to supply side activities.¹² Our chosen list of topics and categories is as follows:

- **Crisis/Recession:** topic - Economic crisis, topic - Crisis, topic - Recession
- **Labour Market:** topic - Unemployment benefits, topic - jobs, topic - Unemployment, Welfare & unemployment
- **Bankruptcy:** topic - Bankruptcy, topic - foreclosure

¹¹See, for example: Guzman (2011); McLaren and Shanbhogue (2011); Askitas and Zimmermann (2009); Fondeur and Karamé (2013); Carrière-Swallow and Labbé (2013).

¹²This list was inspired by previous research such as Woloszko (2020).

- **Credit, Loans & Personal Finance:** topic - Investment, topic - Mortgage, topic - Interest rate, Credit & lending, Investing
- **Consumption Items & Services:** Food & drink, Vehicle brands, Home & garden, Sports, Autos & vehicles, Grocery & food retailers, Vehicle licensing & registration, Hotels & accommodations
- **News:** Business news, Economy news
- **Housing:** topic - Affordable housing, topic - House price index
- **Business & Industrial Activity:** Construction, consulting & contracting, Business services, Transportation & logistics, manufacturing
- **Health:** Health

The relatively large proportion of search items related to consumption of goods and services reflects the large role of consumption in determining U.S. GDP. Vosen and Schmidt (2011) and Woo and Owen (2019) have shown that similar search items track and predict the UMCI index and private consumption very well, thereby capturing consumer sentiment. Labour topics and categories track the popularity of search terms related to job search and benefits demand which Smith (2016), D'Amuri and Marcucci (2017) and Fondeur and Karamé (2013) have shown to predict the unemployment rate in various countries. Topics related to personal finance and investment may signal wealth effects (Woloszko, 2020), which tend to positively correlate with the business cycle (see figure 27 in appendix). Topics around housing have been shown to be indicative of housing prices (Wu and Brynjolfsson, 2015; Askatas and Zimmermann, 2009). The recession, business news and bankruptcy themed search items typically increase during economic downturns which therefore act as signals of economic distress and recessions (Castelnuovo and Tran, 2017; Chen et al., 2012).

While the selection of our search items is subjective, in general, there is no consensus on how to optimally select search terms for final estimation. Methods proposed in the previous literature can be summarised as: (i) pre-screening through correlation with the target variable as found via Google Correlate (Scott and Varian, 2014; Niesert et al., 2020; Choi and Varian, 2012);¹³ (ii) cross-validation (Ferrara and Simoni, 2019); (iii) use of prior economic intuition where search terms are selected through backward induction (Smith, 2016; Ettredge et al.,

¹³Unfortunately, Google Correlate has suspended updating their databases past 2017.

2005; Askitas and Zimmermann, 2009); and (iv) root terms, which similarly specify a list of search terms through backward induction, but additionally download “suggested” search terms from the Google interface. This serves to broaden the semantic variety of search terms in a semi-automatic way. As methodologies based on pure correlation do not preclude spurious relationships (Scott and Varian, 2014; Niesert et al., 2020; Ferrara and Simoni, 2019), we opt for our (somewhat subjective) selection to best guarantee economically relevant Google Trends.

Since search terms can display seasonality, we deseasonalise all Google Trends by the Loess filter, as recommended by Scott and Varian (2014), which is implemented with the “stl” command in R.¹⁴

Although one of the main benefits of Google Trends is their timely availability, which can be as granular as displaying search popularity minutely for the past hour, the purpose of the empirical application is to showcase the flexibility of the proposed models in taking advantage of the heterogeneous information contained in adding new data sources to traditional macroeconomic data, even with little data processing efforts. Due to the simplicity of obtaining monthly Google Trends information, we sample the Google Trends information at the monthly frequency. Nevertheless, the proposed methodology can easily be extended to update monthly Google Trends with higher frequency search information via bridge methods as in Ferrara and Simoni (2019) or could directly be included in the model via expansion of the covariate matrix.¹⁵

The indicative real-time calendar can be found in Table 1 and has been constructed after the data’s real publication schedule. It comprises a total of 37 nowcast periods which make for an equal number of information sets Ω_t^v for $v = 1, \dots, 37$ which are used to construct nowcasts as explained in Section 4. Google Trends are treated as released prior to any other macro information pertaining to a given month, since as argued, Google Trends information can essentially be continuously sampled.

¹⁴To mitigate against inaccuracy stemming from sampling error, we downloaded the set of Google Trends seven times between 1 August 2021 to 8 August 2021 and took the cross-sectional average. Since we used the same IP address and google-mail account, there might still be some unaccounted measurement errors. However, using topics and categories instead of individual search terms, we observe much lower sampling variance.

¹⁵Due to the already very high-dimensionality of the data set, we retain such extensions for future investigation. Constraining the parameter space via MIDAS sampling might make estimation more feasible.

Releases	Timing	Release	Variable Name	Pub. lag	Transformation	FRED Code
1	First day of month 1	No information available	-	-	-	-
2	Last day of month 1	Google Trends		m	4	-
3	4th Friday month 1	Consumer Sentiment	cons	m	3	UMCSENT
4	Last day of month 1	Fed. funds rate & credit spread	fedfunds & baa	m	3	FEDFUNDS & BAAY10
5	1st bus. day of month 2	Economic Policy Uncertainty Index	uncertainty	m-1	1	USEPUINDXM
6	1st bus. day of month 2	PMI	pmi	m-1	1	-
7	1st Friday of month 2	Employment situation	hours & unrate	m-1	2	AWHNONAG & UNRATE
8	Middle of month 2	CPI	cpi	m-1	2	CPI
9	15th-17th of month 2	Industrial Production	indpro	m-1	2	INDPRO
10	3rd week of month 2	Credit & M2	loans & m2	m-1	2	LOANS & M2
11	Later part of month 2	Housing starts	housst	m-1	1	HOUST
12	Last week of month 2	PCE & PCEPI	pce & pce2	m-1	2	PCE & PCEPI
13	Last day of month 2	Google Trends		m	4	-
14	4th Friday month 2	Consumer Sentiment	cons	m	3	UMCSENT
15	Last day of month 2	Fed. funds rate & credit spread	fedfunds & baa	m	3	FEDFUNDS & BAAY10
16	1st bus. day of month 3	Economic Policy Uncertainty Index	uncertainty	m-1	1	USEPUINDXM
17	1st bus. day of month 3	PMI	pmi	m-1	1	-
18	1st bus. day of month 3	Construction starts	construction	m-2	1	TTLCONS
19	1st Friday of month 3	Employment situation	hours & unrate	m-1	2	AWHNONAG & UNRATE
20	Middle of month 3	CPI	cpi	m-1	2	CPI
21	15th-17th of month 3	Industrial Production	indpro	m-1	2	INDPRO
22	3rd week of month 3	Credit & M2	loans & m2	m-1	2	LOANS & M2
23	Later part of month 3	Housing starts	housst	m-1	1	HOUST
24	Last week of month 3	PCE & PCEPI	pce & pce2	m-1	2	PCE & PCEPI
25	Last day of month 3	Google Trends		m	4	-
26	4th Friday month 3	Consumer Sentiment	cons	m	3	UMCSENT
27	Last day of month 3	Fed. funds rate & credit spread	fedfunds & baa	m	3	FEDFUNDS & BAAY10
28	1st bus. day of month 4	Economic Policy Uncertainty Index	uncertainty	m-1	1	USEPUINDXM
29	1st bus. day of month 4	PMI	pmi	m-1	1	-
30	1st bus. day of month 4	Construction starts	construction	m-2	1	TTLCONS
31	1st Friday of month 4	Employment situation	hours & unrate	m-1	2	AWHNONAG & UNRATE
32	Middle of month 4	CPI	cpi	m-1	2	CPI
33	15th-17th of month 4	Industrial Production	indpro	m-1	2	INDPRO
34	3rd week of month 4	Credit & M2	loans & m2	m-1	2	LOANS & M2
35	Later part of month 4	Housing starts	housst	m-1	1	HOUST
36	Last week of month 4	PCE & PCEPI	pce & pce2	m-1	2	PCE & PCEPI
37	Later part of month 5	Housing starts	housst	m-2	1	HOUST

Table 1: Real time calendar based on actual publication dates. Transformation: 1 = monthly change, 2 = monthly growth rate, 3 = no change, 4 = LOESS decomposition. Pub. lag: m = refers to data for the given month within the reference period, m-1 = refers to data with a months' lag to publication in the reference period, m-2 = refers to data with 2 months' lag to publication in the reference period.

3.4 Understanding Google Trends

To get a visual understanding of the Google search information, we plot in figure (1) the first 3 principal components of the U-MIDAS transformed Google Trends information. Figure (2) shows the factor loadings.

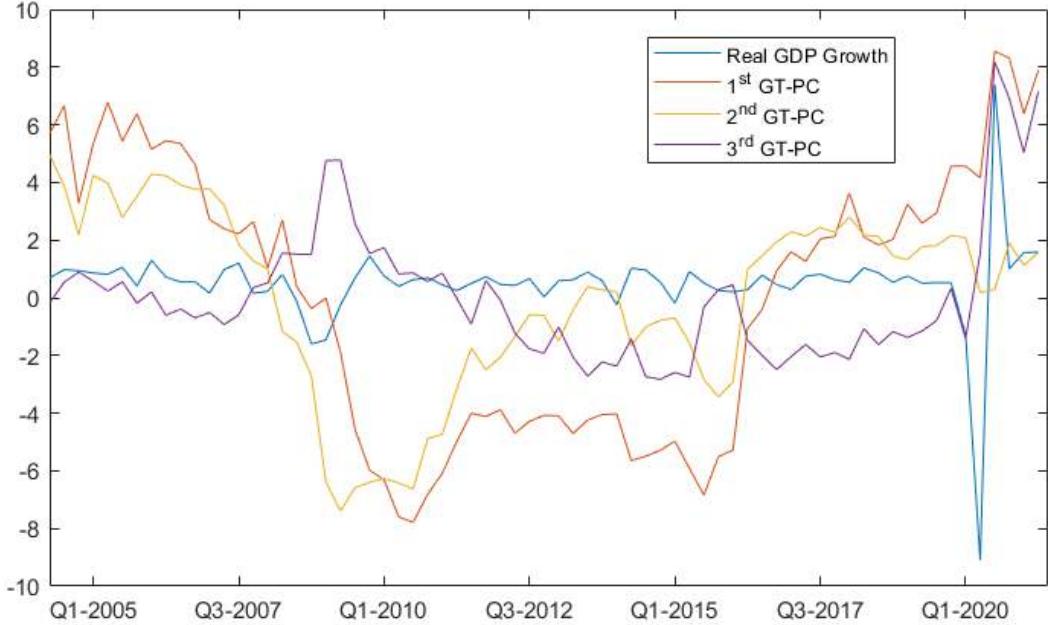


Figure 1: First 3 principal components of the GT data set and 2nd vintage real GDP growth.

The first three principal components show very heterogeneous behaviour, but the dynamics conform to the economic intuition suggested from the loadings. The first component loads positively on supply side activity such as ‘Business services’, ‘Construction, consulting & contracting’, ‘Manufacturing’ as well as ‘Investing’ and negatively on recessionary themes and ‘Jobs’ which spike during the crises and decrease during recoveries (see figure 27 for indicative time-series plots of individual Google search items). Accordingly, the first component decreases strongly during the financial crisis (and to a smaller degree also during height of the Covid-19 recession) and picks up the rapid increase in economic activity after Q2 2020 very well.

The second component can be understood as a measure of consumer sentiment and financial health as it loads mostly on consumption items and negatively on topics such as ‘Bankruptcy’ and ‘Foreclosure’; both terms were very popular during the financial crisis, but not so much the pandemic crisis.¹⁶ Accordingly, the second principal component shows a large dip during

¹⁶This may reflect the different nature of the Covid-19 pandemic induced recession and the positive impact

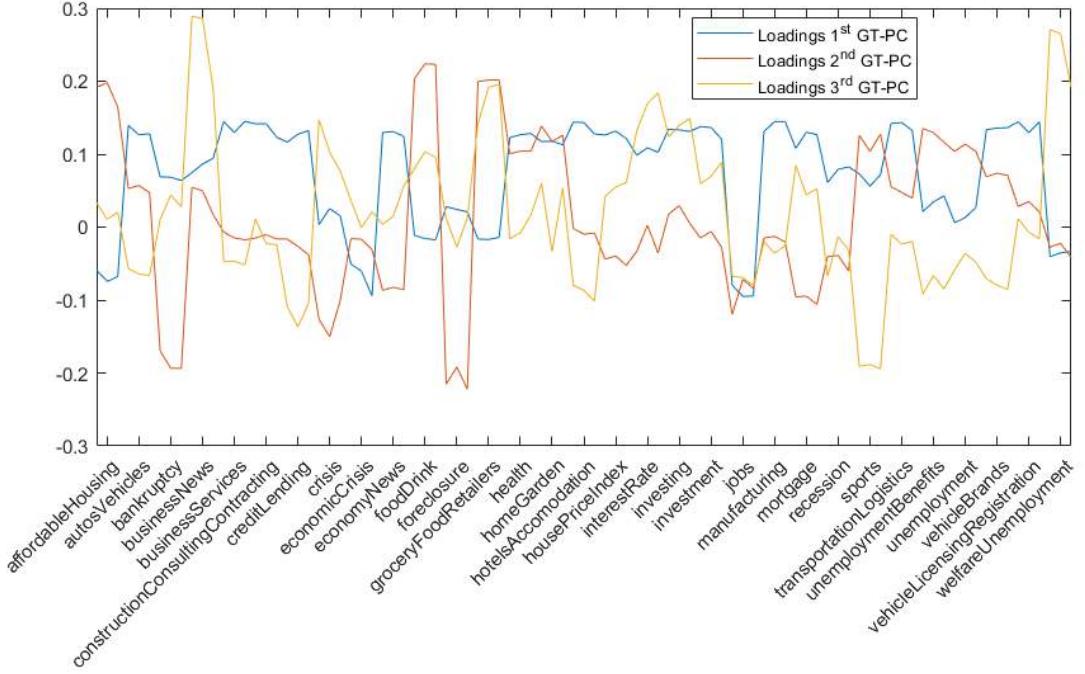


Figure 2: Loadings for first 3 principal components of the GT data set.

the financial but only a minor dip during the Covid-19 recession.

The third principal component loads very strongly on business news/recession/crisis items which increase in popularity during periods characterised by economic anxiety, thus spiking around the financial crisis and the pandemic period. Hence, it can be interpreted as an indicator of economic distress.

3.5 Relationship Between Macroeconomic and Google Search Data

Understanding further how the skip-sampled Google Trend series correlate with the macro data set may help us anticipate which information will likely be picked up in the model. Figure (3) shows a correlation heatmap between the macro data set and the first three Google Trends principal components. Please note that an increase in the UMCI and PMI indicate improved consumer and producer sentiments, respectively.

As expected, the first rather procyclical component correlates highly with the fed-funds rate which tends to also rise with the business cycle. The second component is strongly positively correlated with the UMCI index and negatively with the BAA spread (which increases during deteriorating financial conditions), indicating that it indeed captures something close to con-
government policies had.

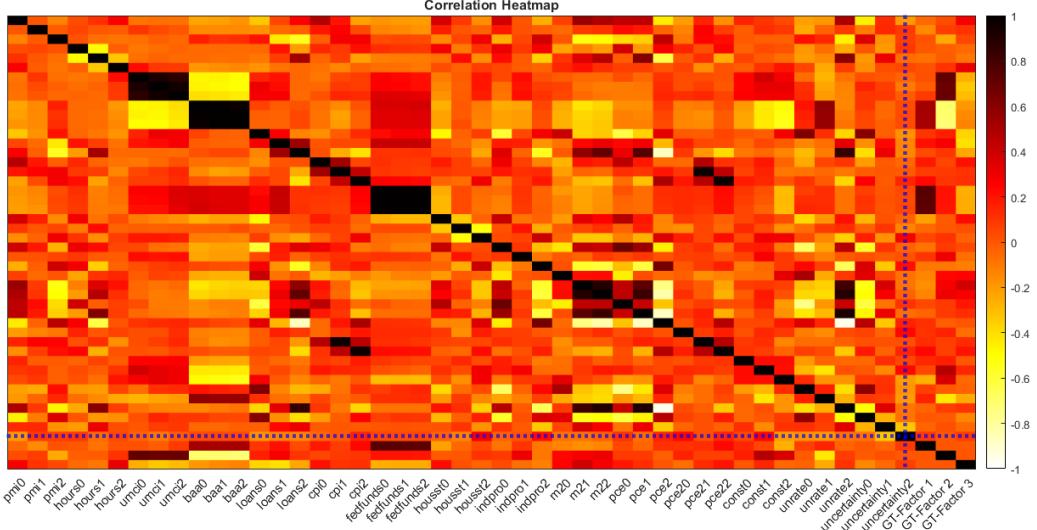


Figure 3: Heatmap of sample correlation matrix.

sumer sentiment and financial health. Since the third component acts as a recession signal, which abruptly spikes during crises, but is otherwise flat, it is not surprising that macroeconomic variables do not correlate very highly with it. This also indicates that search items in this group might add information that is not well captured by the other included macroeconomic information.

4 Nowcasting U.S. Real GDP Growth

The predictive model used to generate in-sample and out-of-sample predictions is given in equations (2)-(3) where the first $T = 45$ observations are used as training sample.¹⁷ We estimate three variants of the model based on priors (7, 9, and 10) and the original BSTS model of Scott and Varian (2014), as well as an AR(4) model for comparison. In line with standard BSTS applications, we first compare the in-sample cumulative absolute one-step-ahead forecast errors, generated from the state space, as well as inclusion probabilities of the variables so as to shed light on which variables produce better fit and explain the outcome. Out-of-sample nowcasts are generated from the posterior predictive distribution $p(y_{T+1}|\Omega_T^v)$ for growth observation y_{T+1} , conditional on the real-time information set Ω_T^v , where ($v = 1, \dots, 37$) refers to nowcast

¹⁷As a further alternative to the proposed BSTS models, we investigate in the appendix as well whether past GDP growth dynamics are more appropriately (in terms of nowcasting) modelled via ARMA components. The results show that the LLT components within the BSTS model are clearly preferred over ARMA type dynamics. We thank an anonymous reviewer for making this suggestion.

periods within the real-time calendar (Table 1). This results in 37 different nowcasts which are generated on a rolling basis until the end of the forecast sample, T_{end} . As recommended by Carriero et al. (2015), variables that have not yet been published until nowcast period v are zeroed out.

Point forecasts are computed as the mean of the posterior predictive distribution and are compared via real time root-mean-squared-forecast-error (RT-RMSFE) which are calculated for each nowcast period as:

$$\text{RT-RMSFE} = \sqrt{\frac{1}{T_{end}} \sum_{j=1}^{T_{end}} (y_{T+j} - \hat{y}_{T+j|\Omega_{T+j-1}^v})^2}, \quad (14)$$

where $\hat{y}_{T+j|\Omega_{T+j-1}^v}$ is the mean of the posterior prediction for nowcast period v using information until $T + j - 1$.

Forecast density fit is measured by the mean real-time log-predictive density score (RT-LPDS) and real-time continuous rank probability score (RT-CRPS):

$$\begin{aligned} \text{RT-LPDS} &= \frac{1}{T_{end}} \sum_{j=1}^{T_{end}} \log p(y_{T+j} | \Omega_{T+j-1}^v) \\ &= \frac{1}{T_{end}} \sum_{j=1}^{T_{end}} \log \int p(y_{T+j} | \Omega_{T+j-1}^v, \zeta_{1:T+j-1}) p(\zeta_{1:T+j-1} | \Omega_{T+j-1}^v) d\zeta_{1:T+j-1} \\ &\approx \frac{1}{T_{end}} \sum_{j=1}^{T_{end}} \log \left(\frac{1}{M} \sum_{m=1}^M p(y_{T+j} | \Omega_{T+j-1}^v, \zeta_{1:T+j-1}^m) \right), \end{aligned} \quad (15)$$

$$\text{RT-LPDS} = \frac{1}{T_{end}} \sum_{j=1}^{T_{end}} \frac{1}{2} \left| y_{T+j} - y_{T+j|\Omega_{T+j-1}^v}^v \right| - \frac{1}{2} \left| y_{T+j|\Omega_{T+j-1}^v}^{v,A} - y_{T+j|\Omega_{T+j-1}^v}^{v,B} \right|, \quad (16)$$

where, for brevity of notation, $\zeta_{1:T+j-1}$ collects all model parameters as defined for each model, which are estimated with expanding in-sample information until $T + j - 1$ and M stands for iterations of the Gibbs sampler after burn-in. Note that in (16), $y_{T+j|\Omega_{T+j-1}^v}^{v,A,B}$ are independently drawn from the posterior predictive density $p(y_{T+1|\Omega_{T+j-1}^v}^v | y_T)$.

As shown by Frühwirth-Schnatter (1995), in a setting where time-varying and fixed components for a structural state space model are chosen, the LPDS can be interpreted as a log-marginal likelihood based on the in-sample information and therefore provides a model founded scoring rule. The RT-CRPS can be thought of as the probabilistic generalisation of the mean-absolute-forecast-error. Similar to the log-score, it belongs to the broader class of

strictly proper scoring rules (Gneiting and Raftery, 2007) which allows for comparing density forecasts in a consistent manner.^{18,19} To facilitate discussion, our objective is to maximise the RT-LPDS and minimise the RT-CRPS. For all forecast metrics, the predictive distribution used for (14, 15, and 16) is traditionally generated in state space models via the prediction equations of the Kalman filter (Harvey, 1990). Instead, we use the simpler approximate method of Cogley et al. (2005), which we found to make no practical difference in our sample.²⁰ The method is described in Appendix A.4.2.

Finally, to test whether a state variance is equal to zero, we make use of the Dickey-Savage density ratio evaluated at $\sigma^{\tau,\alpha} = 0$:

$$DS = \frac{p(\sigma^{\tau,\alpha} = 0)}{p(\sigma^{\tau,\alpha} = 0|y)} \quad (17)$$

It can be shown that for nested models, the DS statistic is equivalent to the Bayes factor between the prior and the posterior distribution of the parameter of interest at zero (Verdinelli and Wasserman, 1995). The intuition for the test is simple: if the prior probability-density-function (PDF) allocates more mass at 0 than the posterior at that point, there is evidence in favour of the unrestricted model, i.e., $\sigma^{\tau,\alpha} \neq 0$. While the priors for the state variances have well known forms and thus can be evaluated analytically, we estimate the denominator for all models through Monte Carlo integration.

4.1 In-Sample Results

Figure 4 shows the in-sample cumulative-one-step ahead prediction errors using the proposed priors where the information set pertains to the entire estimation sample without ragged edges. From Figure 4, it is clear that the horseshoe prior BSTS (HS-BSTS) provides the best in-sample predictions at all time periods. The HS-BSTS-SAVS and SSVS-BSTS initially provide similar fit, however diverge in performance around the financial and Covid-19 crises, especially the HS-SAVS-BSTS. It is striking that, compared to the former two, the HS-BSTS provides very

¹⁸We do not report calibration tests, as there are too few out-of-sample observations to meaningfully determine calibration.

¹⁹Although the CRPS is a symmetric scoring rule, it penalises outliers less aggressively than the log-score which is of advantage in small forecast samples such as ours.

²⁰The Kalman filter provides conditionally optimal forecast densities in terms of squared forecast error. However, if there is misspecification or if the forecast horizon is very short, then approximate methods can do just as well empirically. A similar logic holds when comparing direct and iterative forecasting methods such as in Marcellino et al. (2006). We thank an anonymous reviewer for bringing this to our attention.

stable performance as indicated by a nearly linear increase in errors even during the financial crisis and the Covid pandemic. It is also apparent that the SAVS algorithm is not able to retain the fit of the HS prior alone, which, as we show in the next subsection, is in contrast to the out-of-sample results.

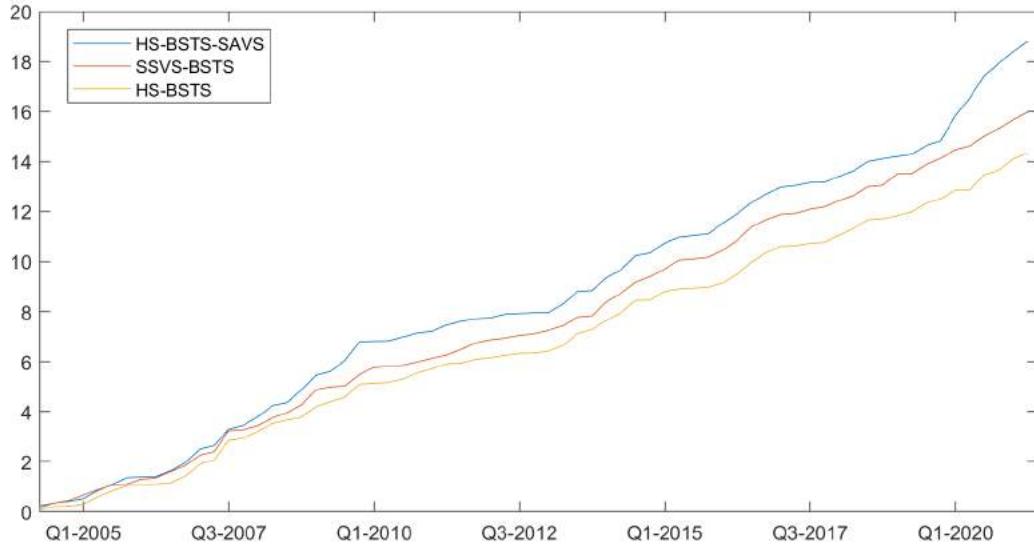


Figure 4: Cumulative one-step-ahead forecast errors in-sample from 3 different models: (1) SSVS-BSTS, (2) HS-BSTS and (3) HS-BSTS-SAIS

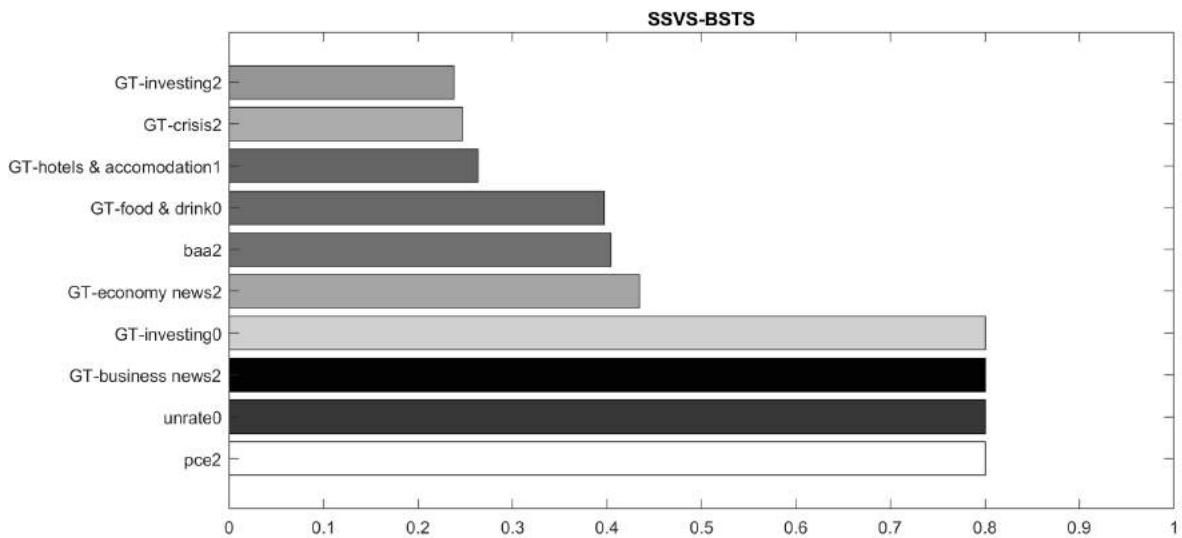


Figure 5: Posterior inclusion probabilities of the SSVS-BSTS model.

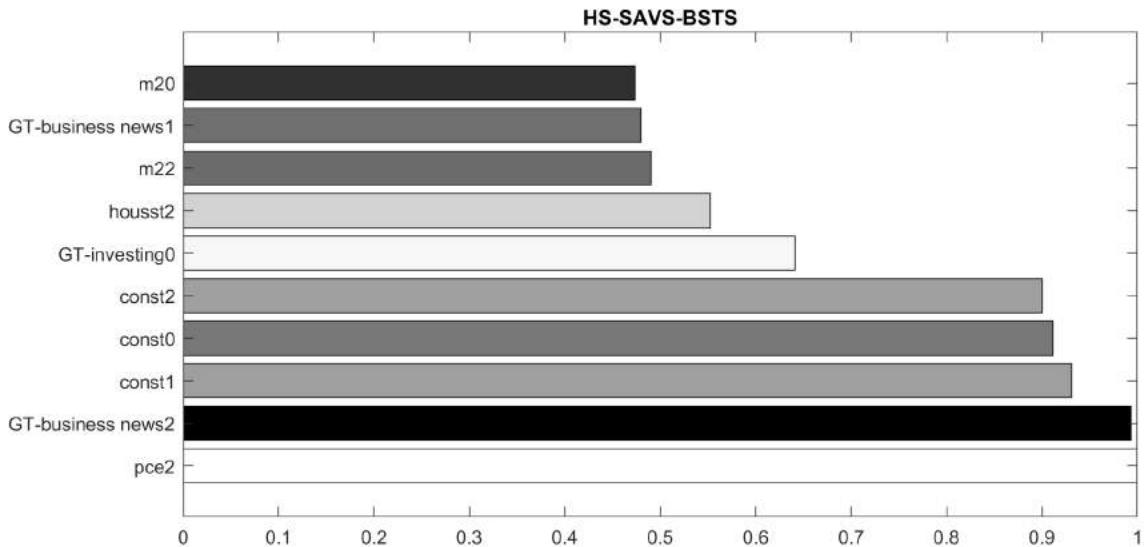


Figure 6: Posterior inclusion probabilities of the HS-SAVS-BSTS model.

To understand the driving variables behind the posterior predictive distributions, the posterior marginal inclusion probabilities for the SSVS and SAVS model are plotted for the top ten most drawn variables in Figures 5 and 6 respectively. The colors of the bars indicate the sign on a continuous scale of white (positive relationship) to black (negative relationship) of the variable when included in the model, and the prefix ‘GT’ indicates Google Trend variables. The number [0,1,2] appended to a variable indicates the temporal position within a given quarter, with 0 being the latest month.

For all three models (for the BSTS model, see Figure 25), the posterior inclusion probabilities show that the most drawn Google search information pertains to the category ‘business news’ and topic ‘investing’ with clearly negative and positive impact respectively on GDP growth. The posteriors on these Google Trends variables (Figure 26) show that not only is the impact statistically significant but also economically so. As search intensity for business news goes up (down), GDP growth forecasts are adjusted downwards (upwards). Vice versa for the investing topic. Since the ‘business news’ category spikes in recessions, but is otherwise flat, this suggests that during periods of heightened recessionary probability and economic fear, people engage and search more for business news which therefore acts as an indicator of expected economic distress. Similar reasoning has led to a large literature on using economic sentiment extracted from news media to model and forecast economic activity (Kalamara and Kapetanios, 2019; Kalamara et al., 2022; Baker et al., 2016; Aprigliano et al., 2022; Gentzkow et al., 2019; Alexopoulos and

Cohen, 2015; Manela and Moreira, 2017; Nyman et al., 2021; Shapiro et al., 2020).²¹

Conversely, ‘investing’ items are presumably searched more often when households and individuals are financially more stable which is when they engage in looking for investment opportunities. As seen in Figure 27, this series tends to positively correlate with the business cycle. This interpretation of the investment topic corroborates findings of Woloszko (2020) who show that the investment topic has a positive impact in a panel data nowcasting exercise for GDP growth.

Figures 5 and 6 also reveal some interesting patterns about how macroeconomic data are employed in the models. The SSVS prior tends to select only the most dominant of the skip-sampled information, while the SAVS extended HS prior allocates significant inclusion probability to all months within a quarter. For example, while the SSVS prior selects the variable ‘unrate0’, i.e., the unemployment rate for the last month in a given quarter, the HS-SAVS prior allocates nearly the same inclusion probability to data for all months on construction starts and M2. This result is likely driven by the fact that the SSVS prior discretises the model space and therefore, with correlated data, will tend to include only the variable with the highest marginal likelihood. Continuous shrinkage priors, on the other hand, make use of all covariates since they are always included in the model.

In line with this discussion, the posterior inclusion probability heatmaps in Figures 7 and 8 show generally that the SSVS and HS prior also display very different degrees of model uncertainty. Note that in the figures, inclusion probabilities to the left of the dashed line pertain to the macroeconomic data set, and to the right, the Google search data. The HS prior tends to display substantial uncertainty over inclusion, particularly for the Google Trends data, which makes sense given the similarity in signal within Google Trends categories and topics. By contrast, the SSVS tends to load only on a few Google search items and not explore posteriors of correlated GTs.

Nonetheless, the fact that the HS prior identifies individual macroeconomic series such as ‘pce2’ as a clear signal indicates that mixed frequency information matter for predictive purposes, which would otherwise be lost in averaging information across quarters.

These results contribute to the recently popularised studies of sparsity within economic prediction problems (Giannone et al., 2021; Cross et al., 2020) in at least two ways. Firstly, they indicate that different sparsity patterns can emerge within data sources (here, macroeconomic

²¹It would be interesting for future research to investigate whether Google Trends and news sentiment extracted from articles substitute or complement each other in modelling recessionary risks.

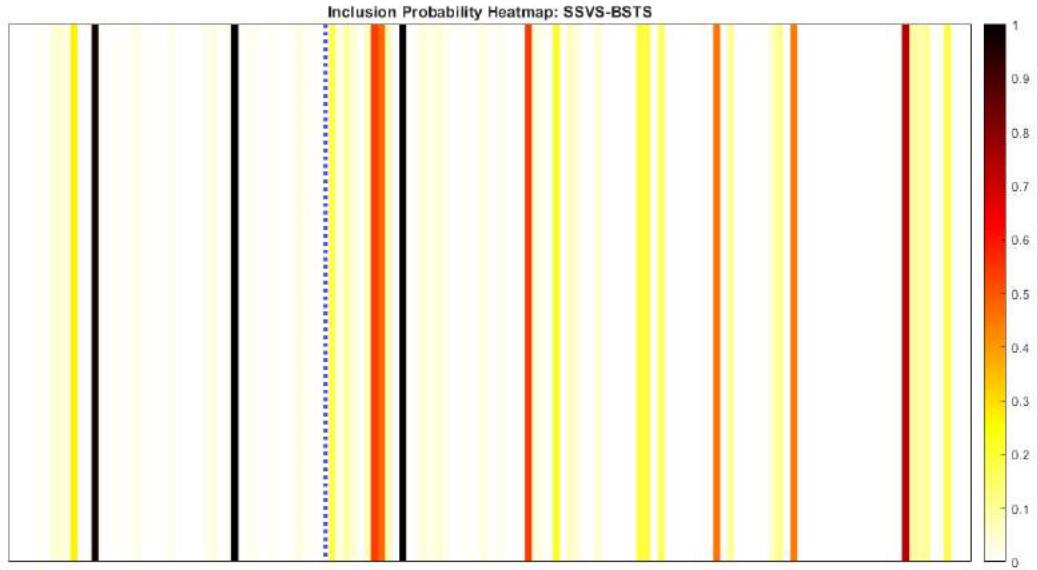


Figure 7: Posterior Inclusion probability heatmaps for the SSVS-BSTS. Inclusion probabilities to the left of the dashed line pertain to the macro data set, and to the right, the Google search data.

and Google Trends) and within mixed frequency information. And secondly, different sparsity patterns can emerge depending on the prior used. Section 5 further investigates robustness of the proposed priors to different sparsity settings, and provide recommendations.

Finally, our in-sample results clearly demonstrate (Figure 9) that there is support in the data for a local trend, but not a local linear trend: the posterior for σ^τ is clearly bi-modal with less mass on zero than the prior, while the posterior for σ_α has substantially more mass on zero than the prior. The Bayes factors are 28.69 and 0.42 for the state standard deviations respectively.

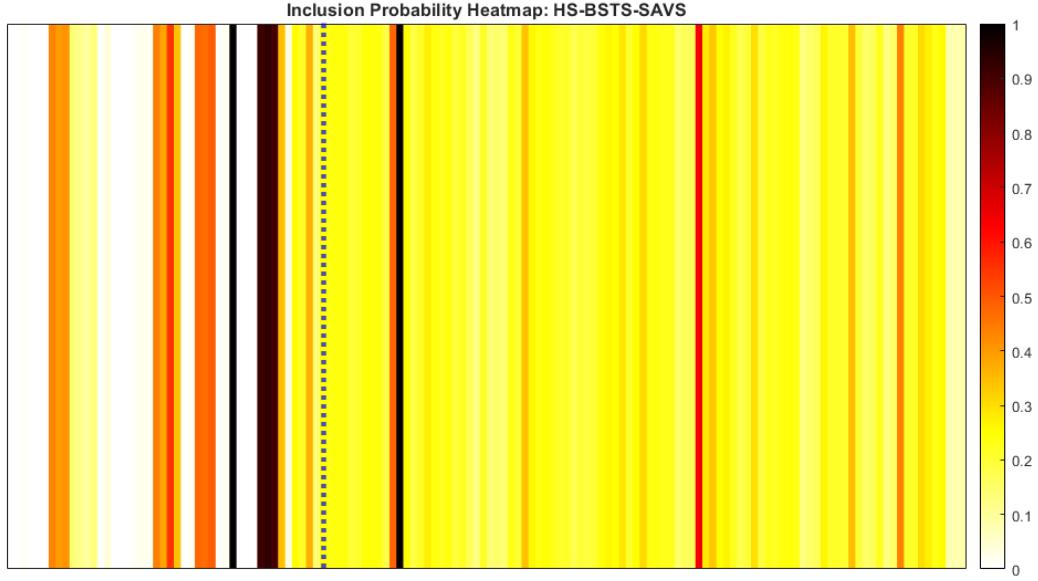


Figure 8: Posterior Inclusion probability heatmaps for the HS-SAVS-BSTS. Inclusion probabilities to the left of the dashed line pertain to the macro data set, and to the right, the Google search data.

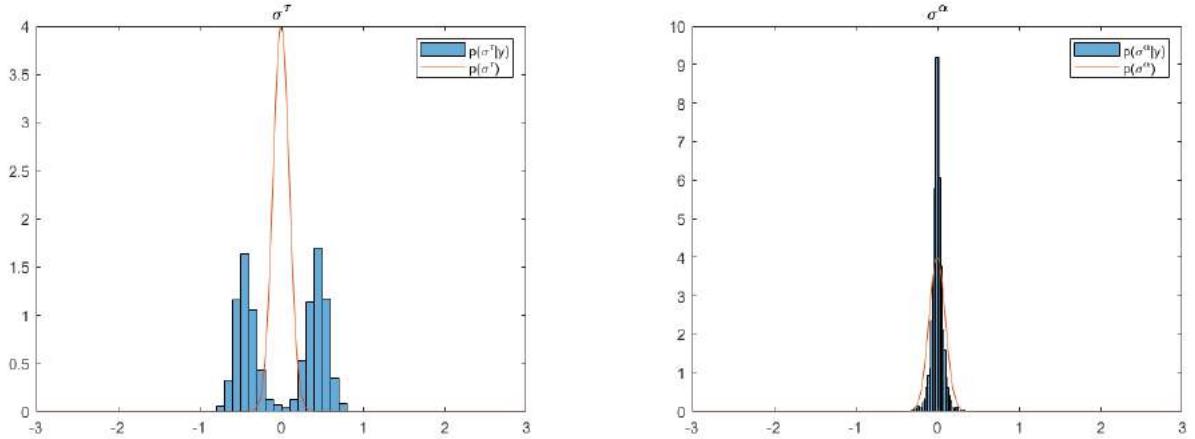


Figure 9: Distribution of the (left) trend state standard deviation and (right) slope standard deviation for the HS-BSTS model.

4.2 Nowcast Evaluation: Pre-Pandemic period

We now turn to out-of-sample nowcasting performance, where nowcasts are produced following the real time data publication calendar as explained in Section 2. Due to the extraordinary

economic circumstances of the Covid-19 pandemic, we split the evaluation sample into: (a) pre-Covid (ending Q4 2019); and (b) during Covid (ending Q2 2021). We first evaluate point-and then density fit for the pre-Covid period.

RT-RMSFE are plotted in Figure 10 for the competing non-centred BSTS estimators, as well as the AR(4) benchmark and the original BSTS model. Note that in all nowcast figures, we represent nowcast periods in which Google Trends are published by grey vertical bars. The following points emerge from Figure 10. Firstly, it is clear that all proposed BSTS models based on the non-centred state space offer large performance gains (for certain nowcast periods up to 40%) over the original BSTS model. Secondly, all models nearly monotonically increase in precision as more data are released, where, as expected, the BSTS models outperform the AR benchmark²² as soon as the first data becomes available. Thirdly, among non-centred BSTS models, the HS-SAVS-BSTS does the best; however, it is closely followed by the HS-BSTS. This indicates that the SAVS algorithm successfully shuts down contributions of noise variables and thus gives further validity to the variable selection results discussed above. With only a modest decrease of 2-5% in RT-MSFE relative to the plain HS-BSTS model, it is evident, however, that the horseshoe prior already provides aggressive shrinkage. Compared to the SSVS-BSTS, the horseshoe prior based BSTS models offer 7-25% improvements in terms of point forecast accuracy, especially so in the beginning nowcast periods.

Finally, we find that there is a large decrease in point-forecast error due to Google Trends releases prior to macroeconomic data becoming available which is consistent across all models considered. Improvements are in the range of 7-25% for the given models compared to the first period nowcasts.²³ The subsequent value of Google Trends for point forecasts is a function of how much a given model loads on the Google Trends search variables and how well the shrinkage prior can separate signal from noise. Hence, improvements for the HS-BSTS models are modest after the first GT release and the SSVS-BSTS experiences a noticeable improvement of 15% in the final GT release, whereas the original BSTS model of Scott and Varian (2014) becomes less precise with latter GT releases. The explanation is that the original BSTS model generally struggles with the dimensionality of data set which leads to ineffective variable selection and

²²The HS-BSTS and HS-SAVS-BSTS outperform the AR model at all nowcast periods significantly as measured by the Diebold et al. (1998) test at conventional significance levels. The SSVS-BSTS model does so only with the second nowcast period.

²³Strictly speaking, these nowcasts are also forecasts due to the information set containing only information from the previous quarter.

consequently poor nowcasting performance.

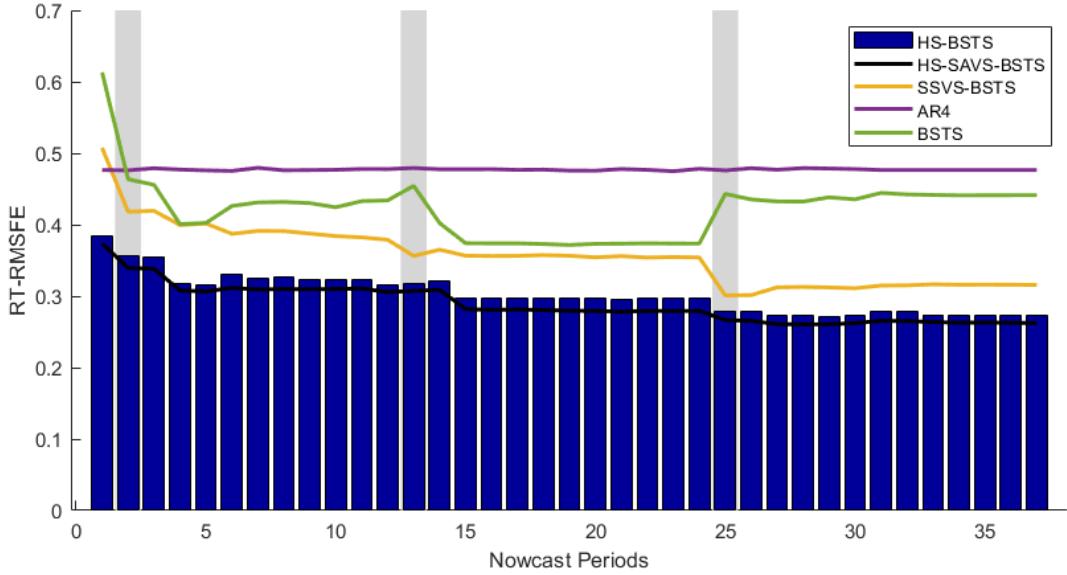


Figure 10: Real-Time RMSFE of all competing models. The RT-RMSFE for the BSTS are plotted on the right axis. Grey vertical bars indicate nowcast periods in which GT are published.

Similar to the real-time point forecasts, we plot real-time LPDS (RT-LPDS) and CRPS (RT-CRPS) in Figures 11 and 12, respectively. The RT-LPDS and RT-CRPS mostly confirm the main findings from the point nowcasts. In contrast to the point nowcasts, however, there is now a much more clear cut performance improvement in density fit when the Google Trends information are released in periods 15 and 27, especially so for the BSTS model of Scott and Varian (2014). This divergence in performance hints at the fact that part of the value of including Google Trends information is to better characterise forecast uncertainty which in turn aids density calibration. Next, we explore this distinctive feature of Google Trends releases for nowcasts during the pandemic.

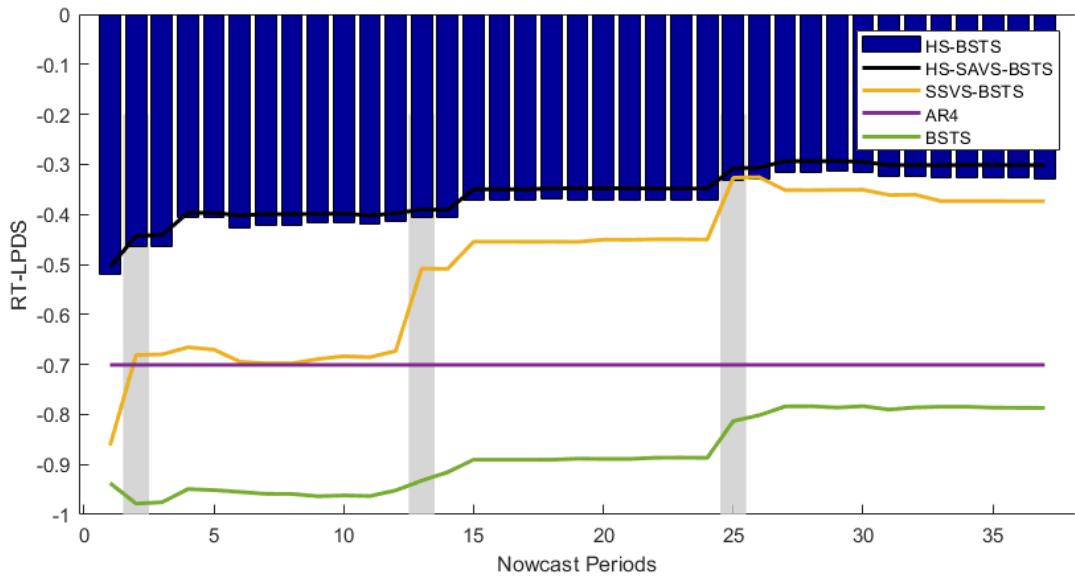


Figure 11: Real-Time log-predictive density scores (RT-LPDS) for all competing models. The RT-LPDS for the BSTS are plotted on the right axis. Grey vertical bars indicate nowcast periods in which GT are published.

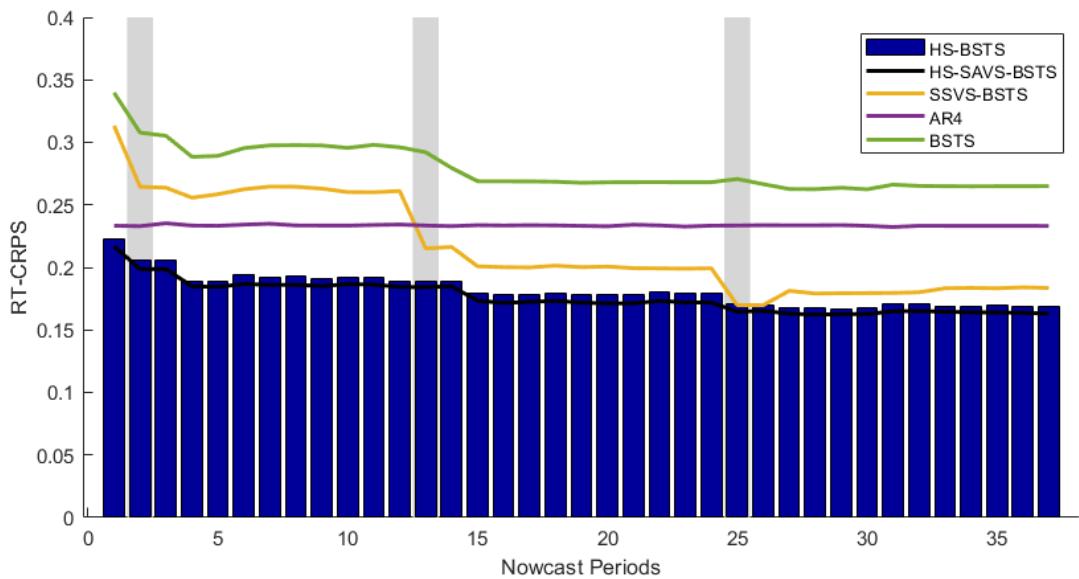


Figure 12: Real-Time CRPS scores (RT-CRPS) for all competing models. The RT-CRPS for the BSTS are plotted on the right axis. Grey vertical bars indicate nowcast periods in which GT are published.

4.3 Nowcast Evaluation: During the Pandemic

The pre-Covid results highlighted that the value of Google Trends are largest before any macroeconomic information are available for the given quarter. While the aim of the nowcasting application for the proposed models is not to provide very granular (weekly or higher) nowcasting models,^{24,25} we now illustrate briefly how, even at the relatively coarse monthly level, Google search information improves predictions during the pandemic.

Figure 13 plots mean forecasts with their credible 95% intervals for the HS-BSTS model based on only macroeconomic data (left column) and the full data set (right column) for the 15th (upper row) and 27th (lower row) nowcast periods respectively. The nowcast periods were chosen to showcase the best possible nowcasts based on information from the end of the second and third month within a quarter respectively. While neither model is able to capture the full extent of the trough during the Covid-19 recession, the model including search term information provides a clear sense of heightened downside risks through a large asymmetric dip of the lower part of credible forecast interval. This is in line with the findings from Woloszko (2020) that for many OECD countries, Google Trends information is able to provide timely downside risk indications. Uniquely, our nowcast exercise highlights how Google Trends can indicate large downward swings in GDP growth over and beyond contributions from macroeconomic data. The fact that Google search information has a greater impact on forecast uncertainty rather than point forecasts further indicates that future research should investigate the potential benefits of using alternative data sources for modelling conditional heteroskedasticity such as in GARCH or stochastic volatility type models.

4.4 An Extension to Student-t Errors

It is also clear that both models struggle to nowcast the equally large upswing that follows the pandemic trough. Inspired by recent VAR forecasting literature during the pandemic (Lenza and Primiceri, 2020; Carriero et al., 2021), we also explore a new BSTS model based on fatter tailed t-distributed errors. The logic behind models with fat tails (compared to those of the normal distribution) is to acknowledge that the large macroeconomic fluctuations, for example

²⁴Higher frequencies would expand the covariate set within the U-MIDAS sampling framework even further. At higher frequencies, predictions could instead be based on single covariates and combined, for example via Bayesian model averaging, or alternatively the frequency weights could be constrained via lower parametric basis functions which is akin to conventional MIDAS estimation.

²⁵See Woloszko (2020) for a panel data approach to a weekly GDP index based on Google Trends.

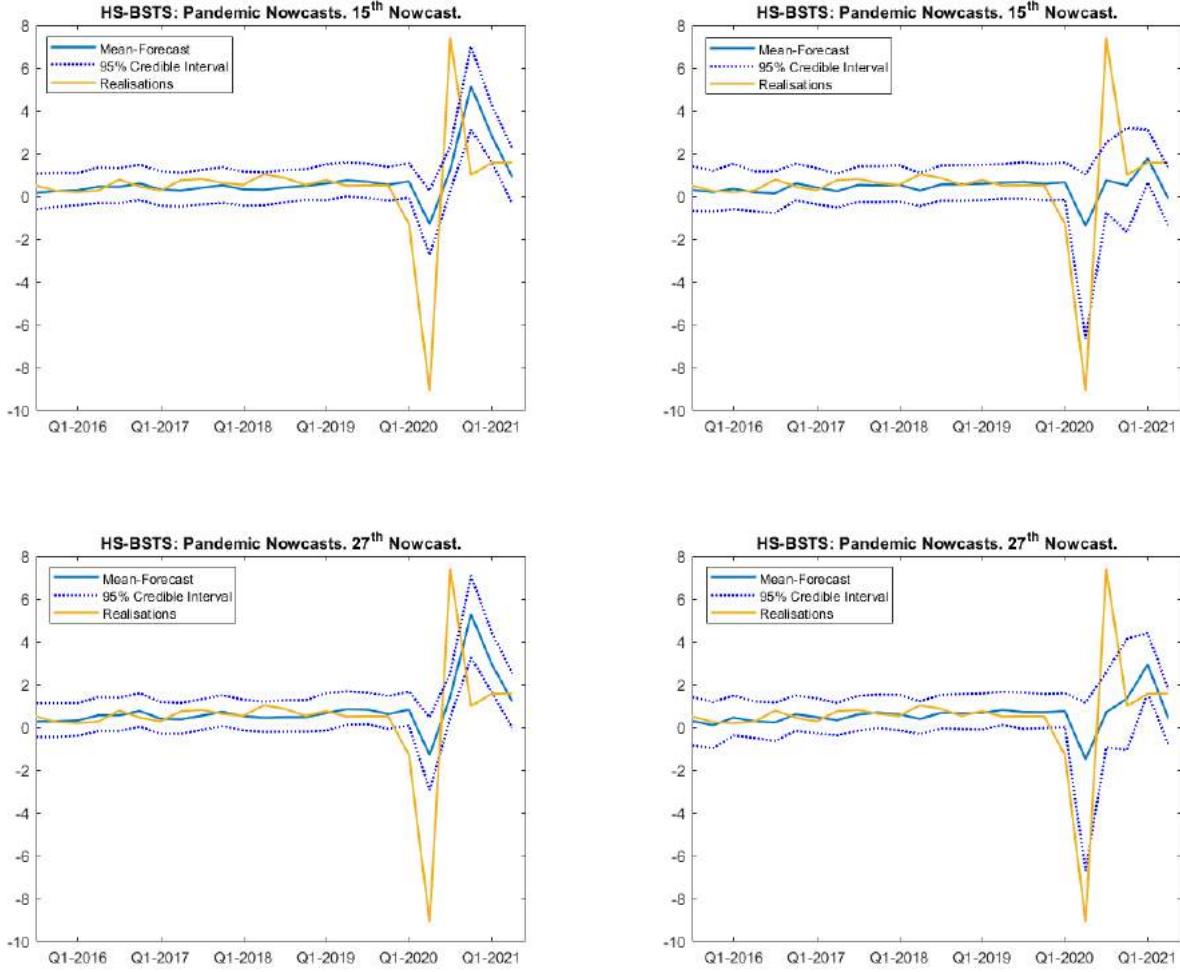


Figure 13: Predictive distributions for the macro only data set (left column) and the macro + Google Trends data set (right columns) for the 15th (upper row) and 27th (lower row) nowcast period.

during the Covid pandemic, are hard to forecast and thus should be modelled through increased forecast uncertainty such that, importantly, large outliers do not adversely affect inference on model parameters.²⁶ Statistically, this is achieved in the posterior by down-weighting outliers through the error covariances.

The estimated model uses the same regression and state components as model (2)-(3). However, we assume $\epsilon_t \sim N(0, \sigma^2 \psi_t)$, where ψ_t is distributed as $\mathcal{G}^{-1}(\eta/2, \eta/2)$ where \mathcal{G}^{-1} denotes the inverse-Gamma distribution and η the degree of freedom parameter of the t-distribution.

²⁶This assumes that the outlier represents an ‘irregular’ observation due to a shock rather than the co-evolution of macroeconomic variables.

Smaller degrees of freedom indicate fatter tails. To estimate the BSTS-t model, we leverage a mixture representation of the t-distribution for which derivations and sampling steps are detailed in appendix (A.5). Treating η as a random variable, figure (23), based on the whole estimation sample shows that there is clear evidence for fatter tails, which are mostly due to the large outliers during the Covid pandemic.

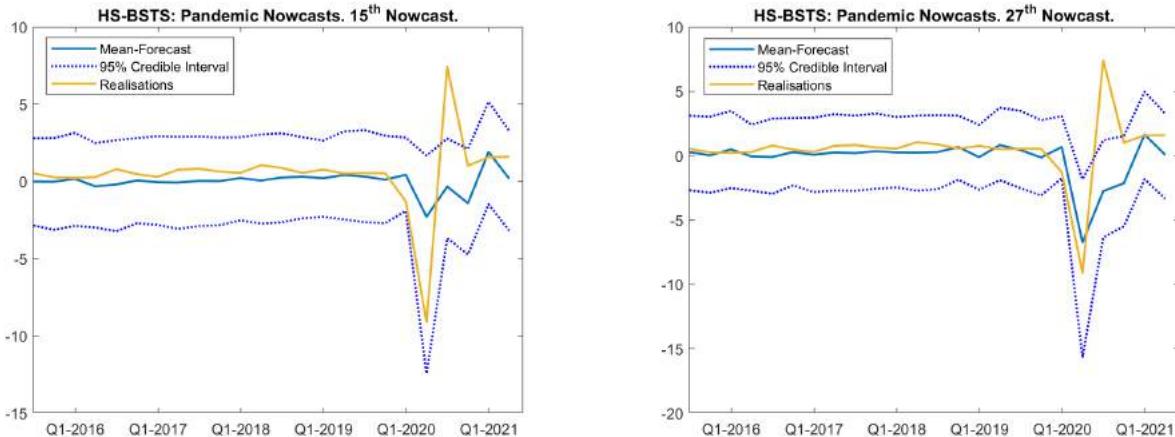


Figure 14: Predictive distributions for the full HS-BSTS-t model.

The nowcasts from this model (Figure 14) show that, in line with the finding of small posterior degrees of freedom for the error distribution, the forecast intervals are much wider compared to the normal BSTS models. The lower forecast interval now captures the trough during the pandemic already in the 15th nowcasting period. Surprisingly, we find that the t-model's mean prediction comes much closer to GDP growth realisation at the height of the recession in period 27. In fact, the additional nowcast period forecasts in Figure (24) show that already with the third publication of the Google search information within Q2 2020, we see a large downward adjustment which had not materialised in period 24 before the GT release. The posterior inclusion probabilities (22) in the appendix reveal that this is because the model loads less heavily on PCE inflation of the first month ('pce2') which had not reacted much until Q2 2020, as opposed to the Google Trends and construction starts data. Yet, this new model still struggles with the upswing and presents the trade-off that even pre-Covid the predictive uncertainty is very large compared to the normal BSTS models. We believe that future research may investigate whether and how different data and modelling techniques are able to accurately forecast not only the trough, but also the peak after sharp downturns.

5 Simulation Study

The empirical application above showed that the proposed BSTS models perform better in point as well as density forecasts compared to the original model of Scott and Varian (2014) and that both the SAVS augmented horseshoe prior as well as the SSVS-BSTS exhibit a relatively sparse selection of macroeconomic data. This finding is in contrast to previous studies using macroeconomic data such as Giannone et al. (2021) and Cross et al. (2020) who find that priors yielding dense models generally outperform sparsity favouring priors. Since an innovation in this paper is the estimation of a latent local-linear trend which might filter out co-movement in the macroeconomic data, we compare the ability of the proposed priors to the original BSTS model (Scott and Varian, 2014) in capturing both sparse and dense environments. Further, to make the simulations closer to our empirical application, we additionally test the priors' ability to detect zero state variances.

Specifically, we simulate local-linear-trend models as (2)-(3) having either the trend variance or the local trend variance set to zero, both equal to zero, or neither equal to zero. Accordingly, we generate 20 simulated samples for $(\sigma^\tau, \sigma^\alpha) = \{(0.5, 0), (0, 0.5), (0, 0), (0.5, 0.5)\}$ together with either a dense or a sparse DGP, where the sparse coefficient vector is set to

$$\beta_{sparse} = (1, 1/2, 1/3, 1/4, 1/5, 0_{K-5})' \quad (18)$$

and the dense coefficient vector is

$$\beta_{dense} = \begin{cases} 1/3 & \text{with probability } p_d \\ 0 & \text{with probability } 1 - p_d \end{cases}, \quad (19)$$

where p_d is set to 2/3. For both coefficient vectors, the dimensionality, K , is set to 300 which is high dimensional compared to the number of observations $T = 150$. We account explicitly for mixed frequencies by first generating the covariate matrix according to a multivariate normal distribution with mean 0 and a covariance matrix with its $(i, j)^{th}$ element defined as $0.5^{|i-j|}$ and then skip-sample each covariate individually after the U-MIDAS methodology as in (13). In the simulations, the true regression coefficient values as well as state variances are known; hence, we compare the performance of the different priors via coefficient bias for the regression coefficients and Dickey-Savage density ratios evaluated at zero state variances. Bias is calculated as

$$\text{Root Mean Coefficient Bias} = \sqrt{\frac{1}{20} \|\hat{\beta} - \beta\|_2^2}, \quad (20)$$

where $\hat{\beta}$ refers to the mean of the posterior distribution. We estimate the original BSTS model with the expected model size, π_0 , equal to the true number of non-zero coefficients.

		Sparse				Dense			
$(\sigma^\tau, \sigma^\alpha)$		(0.5,0)	(0.05)	(0.5,0.5)	(0,0)	(0.5,0)	(0,0.5)	(0.5,0.5)	(0,0)
Bias				Bias					
HS	0.034	0.036	0.036	0.034	0.293	0.289	0.289	0.281	
HS-SAVS	0.035	0.035	0.035	0.035	0.33	0.327	0.32	0.321	
SSVS	0.035	0.038	0.036	0.036	0.415	0.416	0.421	0.418	
BSTS	0.02	0.02	0.021	0.021	0.795	0.567	0.582	0.579	
$DS(\sigma^\tau = 0)$				$DS(\sigma^\tau = 0)$					
HS	516.78	0.81	4.267	0.891	1.959	0.701	3.521	1.493	
SSVS	629.41	0.824	0.89	0.19	10.775	3.053	3.804	1.402	
$DS(\sigma^\alpha = 0)$				$DS(\sigma^\alpha = 0)$					
HS	0.062	41.587	722.319	0.026	0.112	1772.907	96.015	0.068	
SSVS	0.058	4.63E+10	1.05E+08	0.005	0.071	1.29E+10	7.56E+04	0.021	

Table 2: Average Dickey-Savage Density ratio and bias results the simulations. Since the SAVS algorithm is performed on an iteration basis after inference, the posterior of $\sigma^{\tau,\alpha}$ remains unaffected, hence receives the same results as the HS-BSTS model.

As can be seen from Table 2, both the non-centred BSTS models as well as the original BSTS model of Scott and Varian (2014) do better in sparse than in dense DGPs which is similar to the finding of Cross et al. (2020). The largest gains of the proposed BSTS models over Scott and Varian (2014) can be found for dense DGPs where the proposed estimators offer gains in estimation accuracy well in excess of 50%. In sparse designs, however, the latter slightly outperforms the former. This is expected given that the spike-and-slab prior uses a point mass prior on zero and that the true expected model size is used. At the same time, it is encouraging that the differences in accuracy are very small. Among the proposed estimators in dense designs, the HS prior BSTS versions are 30-40% more accurate compared to the SSVS-BSTS which is in line with our findings from the empirical application. Hence, these results offer the conclusion that continuous shrinkage priors are clearly preferred over spike-and-slab models in dense DGPs with a latent local-linear trend component.

The Dickey-Savage density ratio tests confirm that the non-centred state space models are able to correctly identify which of the state variances are significant and which are not, even in high dimensional regression settings. However, the test is sensitive to correctly pinning down the regression coefficient vector: in dense designs, where the SSVS prior does worse than the horseshoe prior, the DS tests in cases (0,0.5) and (0,0) show false support for significant σ^τ .²⁷

²⁷Note that we do not report DS tests for the original BSTS model. This is due to the fact that the prior on

6 Conclusion

In this paper, we investigated the added benefit of including a collection of Google Trends (GT) topics and categories in nowcasts of U.S. real GDP growth through the lens of current-generation Bayesian structural time series (BSTS) models. We extended the BSTS of Scott and Varian (2014) to a non-centred formulation which allows shrinkage of state variances to zero in order to avoid overfitting states and therefore let the data speak about the latent structure. We further extended and compared priors used for the regression part which are agnostic about the underlying model dimensions to accommodate both sparse and dense solutions, as well as the widely successful horseshoe prior of Carvalho et al. (2010). To make the posterior of the horseshoe prior interpretable, we applied sparsifying algorithms borrowed from the machine learning literature, which improve upon the excellent fit of the horseshoe prior itself.

We find that Google Trends improve point as well as density nowcasts in real time within the sample under investigation, where largest improvements appear prior to publication of macroeconomic information. This finding is robust across all considered models. The highest posterior inclusion probability for prediction of GDP growth across all models is obtained with the Google topics/categories ‘business news’ and ‘investing’. The time-series dynamics and model impact of these GTs suggest that they provide timely signals of economic anxiety and wealth effects, respectively. Structural implications of this finding may be investigated with larger Google Trend samples and for other countries. The superior performance of the proposed models over the original BSTS model is confirmed in a simulation study which shows that among the proposed models, the horseshoe prior BSTS performs best and the largest gains in estimation accuracy can be expected in dense DGPs. It is further confirmed that the non-centred state priors are able to correctly identify the latent structure, however they are sensitive to the efficacy of the regression prior to detect signals from noise.

Finally, we applied our models to the Covid-19 pandemic period and find that Google Trends information help characterise the uncertainty during the Covid recession and subsequent recovery period. An extension of the BSTS model to student-t errors is also shown to benefit the timeliness of the forecast revisions to the changes in the macroeconomic data.

Our work suggests some important avenues for future research. An aspect which remained

the state variance has no mass on zero and therefore is not testable.

unexplored in this study is that Google Trends might have time varying importance in relationship to the macroeconomic variable under investigation, as highlighted by Koop and Onorante (2019). Search terms can be highly contextual and might therefore be able to predict turning points in some periods but not in others. While, given the limited quarterly observations of Google Trends, our current investigation of this research question is somewhat limited, this will improve in significance over time. Also, nowcasting in contexts where the design is partly dense and partly sparse is a challenging problem. Our work sheds some light on this question, but it also motivates further research in this direction.

References

- Alexopoulos, M. and J. Cohen (2015). The power of print: Uncertainty shocks, markets, and the economy. *International Review of Economics & Finance* 40, 8–28.
- 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.
- Aprigliano, V., S. Emiliozzi, G. Guaitoli, A. Luciani, J. Marcucci, and L. Monteforte (2022). The power of text-based indicators in forecasting the italian economic activity. *International Journal of Forecasting (forthcoming)*.
- Askitas, N. and K. F. Zimmermann (2009). Google econometrics and unemployment forecasting.
- Baker, S. R., N. Bloom, and S. J. Davis (2016). Measuring economic policy uncertainty. *The quarterly journal of economics* 131(4), 1593–1636.
- Ba  nbara, 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.
- Belmonte, M. A., G. Koop, and D. Korobilis (2014). Hierarchical shrinkage in time-varying parameter models. *Journal of Forecasting* 33(1), 80–94.
- 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. Fr  hwirth-Schnatter (2019). Achieving shrinkage in a time-varying parameter model framework. *Journal of Econometrics* 210(1), 75–97.
- Bok, B., D. Caratelli, D. Giannone, A. M. Sbordone, and A. Tambalotti (2018). Macroeconomic nowcasting and forecasting with big data. *Annual Review of Economics* 10, 615–643.
- Carri  re-Swallow, Y. and F. Labb   (2013). Nowcasting with google trends in an emerging market. *Journal of Forecasting* 32(4), 289–298.

- 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.
- 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.
- Castelnuovo, E. and T. D. Tran (2017). Google it up! a google trends-based uncertainty index for the united states and australia. *Economics Letters* 161, 149–153.
- 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 macroeconomics. *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.
- Chen, C. W., R. Gerlach, B. B. Hwang, and M. McAleer (2012). Forecasting value-at-risk using nonlinear regression quantiles and the intra-day range. *International Journal of Forecasting* 28(3), 557–574.
- Choi, H. and H. Varian (2012). Predicting the present with google trends. *Economic record* 88, 2–9.
- Clark, T. E. (2011). Real-time density forecasts from bayesian vector autoregressions with stochastic volatility. *Journal of Business & Economic Statistics* 29(3), 327–341.

- Cogley, T., S. Morozov, and T. J. Sargent (2005). Bayesian fan charts for uk inflation: Forecasting and sources of uncertainty in an evolving monetary system. *Journal of Economic Dynamics and Control* 29(11), 1893–1925.
- Cross, J. L., C. Hou, and A. Poon (2020). Macroeconomic forecasting with large bayesian vars: Global-local priors and the illusion of sparsity. *International Journal of Forecasting*.
- Croushore, D. (2006). Forecasting with real-time macroeconomic data. *Handbook of economic forecasting* 1, 961–982.
- 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.
- D'Amuri, F. and J. Marcucci (2017). The predictive power of google searches in forecasting us unemployment. *International Journal of Forecasting* 33(4), 801–816.
- Ettredge, M., J. Gerdes, and G. Karuga (2005). Using web-based search data to predict macroeconomic statistics. *Communications of the ACM* 48(11), 87–92.
- Ferrara, L. and A. Simoni (2019). When are google data useful to nowcast gdp? an approach via pre-selection and shrinkage.
- Fetzer, T., L. Hensel, J. Hermle, and C. Roth (2020). Coronavirus perceptions and economic anxiety. *The review of economics and statistics*, 1–36.
- Follett, L. and C. Yu (2019). Achieving parsimony in bayesian vector autoregressions with the horseshoe prior. *Econometrics and Statistics* 11, 130–144.
- Fondeur, Y. and F. Karamé (2013). Can google data help predict french youth unemployment? *Economic Modelling* 30, 117–125.
- 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)*, 57–82.
- Frühwirth-Schnatter, S. (1994). Data augmentation and dynamic linear models. *Journal of time series analysis* 15(2), 183–202.
- Frühwirth-Schnatter, S. (1995). Bayesian model discrimination and bayes factors for linear gaussian state space models. *Journal of the Royal Statistical Society: Series B (Methodological)* 57(1), 237–246.
- 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.
- Gentzkow, M., B. Kelly, and M. Taddy (2019). Text as data. *Journal of Economic Literature* 57(3), 535–74.
- George, E. I. and R. E. McCulloch (1993). Variable selection via gibbs sampling. *Journal of the American Statistical Association* 88(423), 881–889.
- Ghosh, J., Y. Li, R. Mitra, et al. (2018). On the use of cauchy prior distributions for bayesian logistic regression. *Bayesian Analysis* 13(2), 359–383.
- Giannone, D., M. Lenza, and G. E. Primiceri (2021). Economic predictions with big data: The illusion of sparsity.
- Giannone, D., F. Monti, and L. Reichlin (2016). Exploiting the monthly data flow in structural forecasting. *Journal of Monetary Economics* 84, 201–215.
- Gneiting, T. and A. E. Raftery (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American statistical Association* 102(477), 359–378.
- Grant, A. L. and J. C. Chan (2017). A bayesian model comparison for trend-cycle decompositions of output. *Journal of Money, Credit and Banking* 49(2-3), 525–552.
- Guzman, G. (2011). Internet search behavior as an economic forecasting tool: The case of inflation expectations. *Journal of economic and social measurement* 36(3), 119–167.
- Harvey, A. C. (1985). Trends and cycles in macroeconomic time series. *Journal of Business & Economic Statistics* 3(3), 216–227.

- Harvey, A. C. (1990). *Forecasting, structural time series models and the Kalman filter*. Cambridge university press.
- Hodrick, R. J. and E. C. Prescott (1997). Postwar us business cycles: an empirical investigation. *Journal of Money, credit, and Banking*, 1–16.
- 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.
- Kalamara, E. and G. Kapetanios (2019). Introducing time variation in machine learning models. *Mimeo*.
- Kalamara, E., A. Turrell, C. Redl, G. Kapetanios, and S. Kapadia (2022). Making text count: economic forecasting using newspaper text. *Journal of Applies Econometrics (forthcoming)*.
- 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.
- Konrath, S., T. Kneib, and L. Fahrmeir (2008). Bayesian regularisation in structured additive regression models for survival data.
- 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 (2020). How to estimate a var after march 2020. Technical report, National Bureau of Economic Research.
- Makalic, E. and D. F. Schmidt (2015). A simple sampler for the horseshoe estimator. *IEEE Signal Processing Letters* 23(1), 179–182.

- Manela, A. and A. Moreira (2017). News implied volatility and disaster concerns. *Journal of Financial Economics* 123(1), 137–162.
- Marcellino, M., J. H. Stock, and M. W. Watson (2006). A comparison of direct and iterated multistep ar methods for forecasting macroeconomic time series. *Journal of econometrics* 135(1-2), 499–526.
- 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.
- McLaren, N. and R. Shanbhogue (2011). Using internet search data as economic indicators. *Bank of England Quarterly Bulletin* (2011), Q2.
- Moran, G. E., V. Ročková, E. I. George, et al. (2018). Variance prior forms for high-dimensional bayesian variable selection. *Bayesian Analysis*, 1091–1119.
- Niesert, R. F., J. A. Oorschot, C. P. Veldhuisen, K. Brons, and R.-J. Lange (2020). Can google search data help predict macroeconomic series? *International Journal of Forecasting* 36(3), 1163–1172.
- Nyman, R., S. Kapadia, and D. Tuckett (2021). News and narratives in financial systems: exploiting big data for systemic risk assessment. *Journal of Economic Dynamics and Control* 127, 104119.
- 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*.
- Scott, S. L. and H. R. Varian (2014). Predicting the present with bayesian structural time series. *International Journal of Mathematical Modelling and Numerical Optimisation* 5(1-2), 4–23.

- Shapiro, A. H., M. Sudhof, and D. J. Wilson (2020). Measuring news sentiment. *Journal of Econometrics*.
- Sims, C. A. (2002). The role of models and probabilities in the monetary policy process. *Brookings Papers on Economic Activity 2002*(2), 1–40.
- Smith, P. (2016). Google’s midas touch: Predicting uk unemployment with internet search data. *Journal of Forecasting* 35(3), 263–284.
- Stock, J. H. and M. W. Watson (2012). Disentangling the channels of the 2007-2009 recession. Technical report, National Bureau of Economic Research.
- Verdinelli, I. and L. Wasserman (1995). Computing bayes factors using a generalization of the savage-dickey density ratio. *Journal of the American Statistical Association* 90(430), 614–618.
- Vosen, S. and T. Schmidt (2011). Forecasting private consumption: survey-based indicators vs. google trends. *Journal of forecasting* 30(6), 565–578.
- Watson, M. W. (1986). Univariate detrending methods with stochastic trends. *Journal of monetary economics* 18(1), 49–75.
- Woloszko, N. (2020). Tracking activity in real time with google trends.
- Woo, J. and A. L. Owen (2019). Forecasting private consumption with google trends data. *Journal of Forecasting* 38(2), 81–91.
- Wu, L. and E. Brynjolfsson (2015). The future of prediction: How google searches foreshadow housing prices and sales. In *Economic analysis of the digital economy*, pp. 89–118. University of Chicago Press.
- Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American statistical association* 101(476), 1418–1429.

A Appendix

A.1 Posteriors

In this section of the appendix, we provide the conditional posterior distributions for the regression parameters which complete the sampling steps for the Gibbs sampler detailed in (2.2.5).

A.1.1 Horseshoe Prior

Starting from model 2 and assuming that the states and state variances have already been drawn in steps 1.-3. in 2.2.5 which is further described in (A.4.1) below. We subtract off τ such that $y - \tau = y^* = X\beta + \epsilon$, $N(0, \sigma^2 I_T)$. Printing the prior here again for convenience:

$$\begin{aligned} \beta_j | \lambda_j, \nu, \sigma &\sim N(0, \lambda_j^2 \nu^2 \sigma^2), \quad j \in 1, \dots, K \\ \lambda_j &\sim C_+(0, 1) \\ \nu &\sim C_+(0, 1) \\ \sigma^2 &\propto \sigma^{-2} d\sigma^2. \end{aligned} \tag{21}$$

Then, by standard calculations (see Bhattacharya et al. (2016)):

$$\begin{aligned} \beta | y^*, \lambda, \nu, \sigma &\sim N(A^{-1}X'y^*, \sigma^2 A^{-1}) \\ A &= (X'X + \Lambda_*^{-1}) \\ \Lambda_* &= \nu^2 \text{diag}(\lambda_1^2, \dots, \lambda_K^2) \end{aligned} \tag{22}$$

Instead of computing the large dimensional inverse A^{-1} , we rely on a data augmentation technique introduced by Bhattacharya et al. (2016). This reduces the computational complexity from $\mathcal{O}(K^3)$ to $\mathcal{O}(T^2K)$. Suppose the posterior is normal $N_K(\mu, \Sigma)$ with

$$\Sigma = (\phi'\phi + D^{-1})^{-1}, \quad \mu = \Sigma\phi'\alpha, \tag{23}$$

where $\alpha \in \mathcal{R}^{T \times 1}$, $\phi \in \mathcal{R}^{T \times K}$ and $D \in \mathcal{R}^{K \times K}$ is symmetric positive definite. Bhattacharya et al. (2016) show that an exact sampling algorithm is given by:

Notice that $\phi = X/\sigma$, $D = \sigma^2 \Lambda_*$ and $\alpha = y^*/\sigma$.

The conditional posterior for the error variance is standard (Makalic and Schmidt, 2015):

$$\sigma^2 | y^*, \beta, \lambda, \nu \sim \mathcal{G}^{-1}((T - K)/2, (y^* - X\beta)'(y^* - X\beta)/2 + \beta'\Lambda_*^{-1}\beta/2), \tag{24}$$

where \mathcal{G}^{-1} denotes the inverse gamma distribution.

Algorithm 1 Fast Horseshoe Sampler

- 1: Sample independently $u \sim N(0, D)$ and $\delta \sim N(0, I_T)$
 - 2: Set $\xi = \Phi u + \delta$
 - 3: Solve $(\Phi D \Phi' + I_T)w = (\alpha - \xi)$
 - 4: Set $\theta = u + D \Phi' w$
-

The posteriors of the scales (ν, λ) are non-standard and need different treatment. We follow Polson et al. (2014) who propose an efficient slice sampler. In particular, define $\eta_j = 1/\lambda_j^2$ and $\mu_j = \beta_j/(\sigma\nu)$. Then, the conditional posterior distribution for η_j takes the following form:

$$p(\eta_j | \nu, \sigma, \mu_j) \propto \exp\left(-\frac{\mu_j^2}{2}\eta_j\right) \frac{1}{1 + \eta_j}. \quad (25)$$

The slice sampler is then implemented as follows:

1. Sample $u_j | \eta_j$ uniformly in the interval $(0, 1/(1 + \eta_j))$
2. Sample $\eta_j | \mu_j, u_j \sim \exp(2/\mu_j^2)$, truncated to have zero probability outside $(0, (1 - u_j)/u_j)$.

Now, transforming back to λ_j yields a direct draw from its posterior distribution and by setting $\eta = 1/\nu^2$ and replacing $\mu_j^2 = \sum \beta_j^2/2$ yields an equivalent draw from the conditional posterior of ν . The advantages of the slice sampling algorithm include: its simplicity; that it involves no rejections; and that it requires no external parameters to be set.

A.1.2 SSVS Prior

Conditioning on the states as in A.1.1, we apply the prior:

$$\begin{aligned} \beta_j | \gamma_j, \delta_j^2 &\sim \gamma_j N(0, \delta_j^2) + (1 - \gamma_j) N(0, c \times \delta_j^2) \\ \delta_j^2 &\sim \mathcal{G}^{-1}(a_1, a_2) \\ \gamma_j &\sim \text{Bernoulli}(\pi_0) \\ \pi_0 &\sim \mathcal{B}(b_1, b_2), \end{aligned} \quad (26)$$

where \mathcal{B} stands for the beta distribution. The conditional posteriors are standard and derived for example in George and McCulloch (1993) and Ishwaran et al. (2005). The difference to the prior of George and McCulloch (1993) lies in the additional prior for δ_j^2 which is assumed to be inverse gamma. It can be shown that this implies a mixture of student-t distributions for β_j marginally (Konrath et al., 2008). We sample from the conditional posteriors in the following way:

Algorithm 2 SSVS Sampler

- 1: For $j \in \{1, \dots, K\}$, sample each $\gamma_j | \beta_j, \delta_j^2, \pi_0, y \sim (1 - \pi)N(\beta_j | 0, c \times \delta_j^2)I_{\gamma_j=0} + \pi_0 N(\beta_j | 0, \delta_j^2)I_{\gamma_j=1}$
 - 2: Sample $\pi_0 \sim \mathcal{B}(b_1 + n_1, b_2 + K - n_1)$, where $n_1 = \sum_j I_{\gamma_j=1}$
 - 3: Sample $\beta | \gamma, \delta^2, \sigma^2, y \sim N(A^{-1}X'y^*/\sigma^2, A)$, where $A^{-1} = X'X/\sigma^2 + D - 1$, $D = \text{diag}(\delta_j^2 \gamma_j)$
 - 4: Sample $\sigma^2 \sim \mathcal{G}^{-1}(\bar{c}, \bar{C})$, where $\bar{c} = \underline{c} + \frac{T}{2}$, $\bar{C} = \underline{C} + \frac{1}{2}((y^* - X\beta)'(y^* - X\beta))$ and $p(\sigma^2) \sim \mathcal{G}(\underline{c}, \underline{C})$
-

A.2 Robustness Check for the Horseshoe Prior

The Cauchy distribution can be a challenging distribution to fit. Due to no analytically available moments, a posterior distribution in which the prior information dominates the likelihood, Cauchy priors might induce vanishing posterior moments (Ghosh et al., 2018; Piironen et al., 2017). Piironen et al. (2017) provide a way in which potential problems due to weak identification can be diagnosed, which is based on the prior induced effective model size distribution.

Assuming a scale mixture of normal prior such as (2.2.3), the conditional posterior $p(\beta | \Lambda_*, \tau, \sigma^2, y)$ within a normal linear regression can be written as:

$$\bar{\beta}_j = (1 - \kappa_j)\hat{\beta}_j \quad (27)$$

where

$$\kappa_j = \frac{1}{1 + T\sigma^{-2}\nu^2 s_j^2 \lambda_j^2}, \quad (28)$$

and $\bar{\beta}_j$ and $\hat{\beta}_j$ refer to the mean of the posterior for $\beta_j | \bullet$ and the maximum likelihood solution respectively. κ_j can be regarded as a shrinkage coefficient as it is defined over the unit interval and therefore dictates how much shrinkage is applied to the maximum likelihood solution. It is easy to verify that when $\lambda_j\nu \rightarrow \infty$, then $\kappa_j \rightarrow 0$ and when $\lambda_j\nu \rightarrow 0$, then $\kappa_j \rightarrow 1$. The distribution of the shrinkage coefficient $p(\kappa_j | \Lambda_*, \nu, \sigma^2, y)$ is implicitly defined through the priors for the hyperparameters (Λ, ν). By applying the change of variable theorem, it can be shown that for the horseshoe prior, this distribution is proportional to $\mathcal{B}(0.5, 0.5)$. In fact, by plotting this distribution alongside the distribution implied by a spike-and-slab prior with zero point-mass spike and slab with infinite scale, the horseshoe prior provides a continuous approximation to the SSVS.

Now, by summing over the shrinkage coefficients in (28), the authors provide a measure of “active regressors”: $m_{eff} = \sum(1 - \kappa_j)$, i.e. large slopes, which are used to illicit a prior for

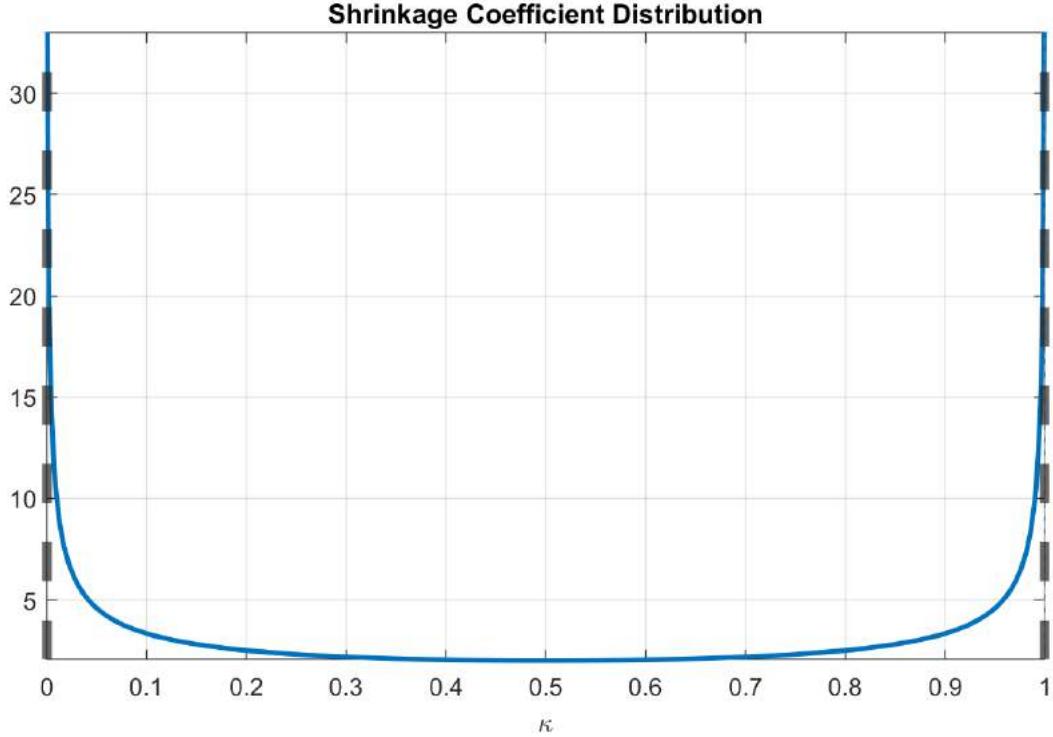


Figure 15: $\kappa|\bullet$ distributions for the horseshoe prior (blue lines) and the spike-and-slab prior (dashed grey lines) with zero point mass and infinite slab scale.

the global shrinkage scale. Due to the aggressive shrinkage profile of the horseshoe prior, the distribution over m_{eff} can be thought of as an effective model size distribution.

Piironen et al. (2017) show that, drawing from the prior predictive distribution with the standard horseshoe hierarchy, the prior effective model size distribution diverges to the largest possible number of slopes. This is confirmed in Figure (16).

Since, as discussed above, the prior predictive distribution of m_{eff} diverges to K, weak identification of either the local or global scale parameters would lead to a divergent posterior m_{eff} distribution, which we use as a diagnostic tool.

However, some caveats must be highlighted: (1) the number of active coefficients formulation is derived from the *conditional* posterior, and hence does not account for uncertainty in (ν, σ^2) ; and (2) it makes further two strong assumptions: the covariates are uncorrelated and the maximum likelihood solution to β exists. Hence, caution should be exercised in taking this approach at face value with high-dimensional macro data, which have typically large cross-sectional correlation.

Nevertheless, we provide in Figure 17 the posterior effective model size distributions for

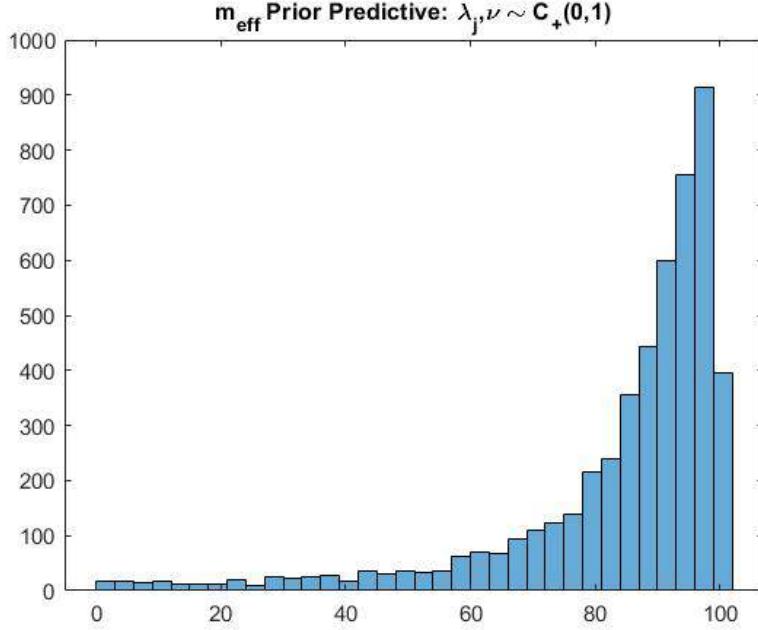


Figure 16: m_{eff} generated by setting $T = K = 100$ and $\sigma = 1$. Distribution based on 5000 samples according to the prior in (2.2.3).

the HS-BSTS model based on the entire sample without ragged edges. As the figure clearly shows, the effective model size is not divergent to K and is very similarly distributed to the SAVS model size distribution. Hence, for the HS-BSTS model, there is no indication of weakly identified posterior scale processes.

A.3 ARMA vs LLT Estimation

While the nowcasting exercise has demonstrated the statistical support for a local-linear-trend model, and hence support for shifts in the long-run rate of GDP growth, it is an important question as to whether GDP dynamics within a high-dimensional regressor setting is better modelled via ARMA components.

Hence, we also estimate the regression as in the paper, but with ARMA components instead of a LLT. The model considered is:

$$y_t = \beta_0 + x'_t \beta + \sum_{j=1}^p \rho_j y_{t-j} + \epsilon_t \quad (29)$$

$$\epsilon_t = u_t + \sum_{m=1}^q \psi_m u_{t-m},$$

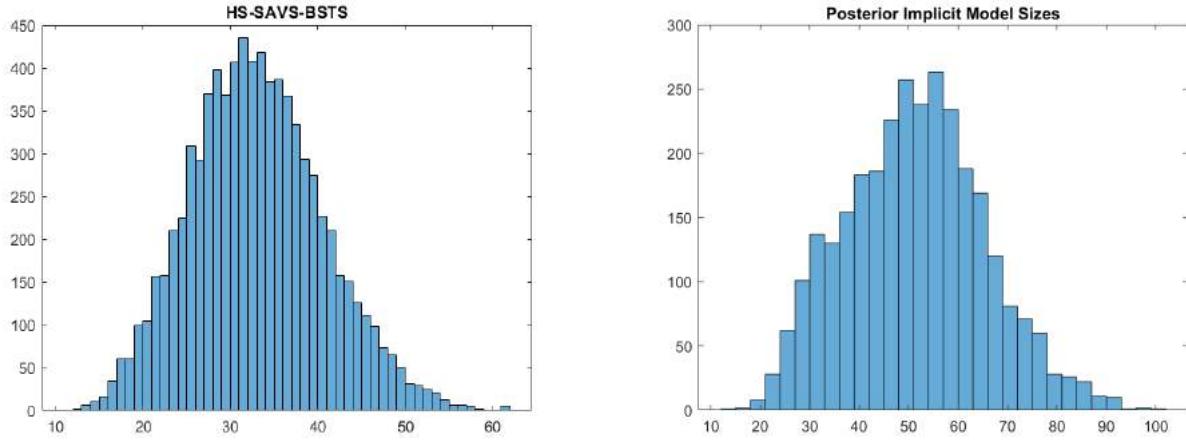


Figure 17: Posterior model size distribution for the HS-SAVS-BSTS (left) and the posterior of m_{eff} for the HS-BSTS (right)

where $|\psi| < 1$ for identification purposes and $\{u_t\}_0^T \sim N(0, \sigma^2)$.

The posterior for the AR coefficients is modelled via the horseshoe prior. We further assume a uniform prior on the interval (-1,1) for $\psi \sim \mathcal{U}(-1, 1)$. To estimate the ARMA components, we follow Chan (2017a) by using a band and sparse matrix representation which allows for very fast computation by avoiding recursive algorithms. The order of the ARMA components are chosen via the best out-of-sample performance.

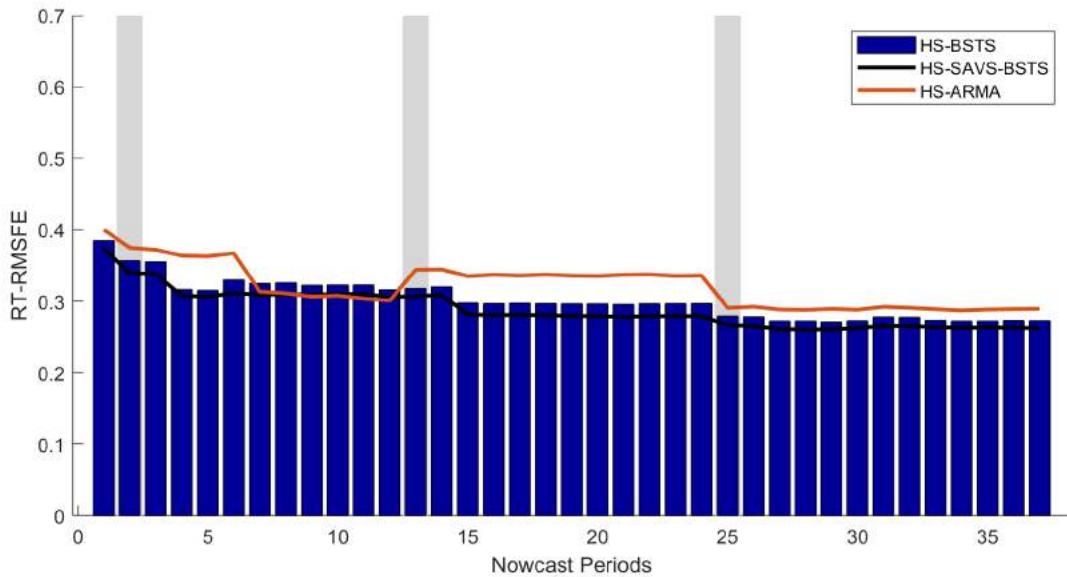


Figure 18: Real-Time-RMSFE for 37 nowcast periods.

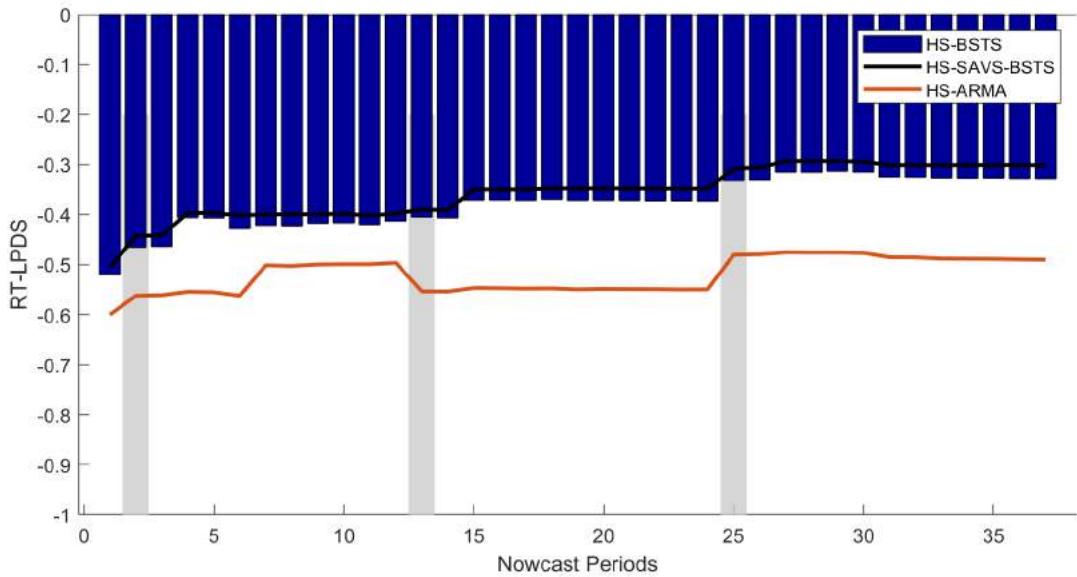


Figure 19: Real-Time-LPDS for 37 nowcast periods.

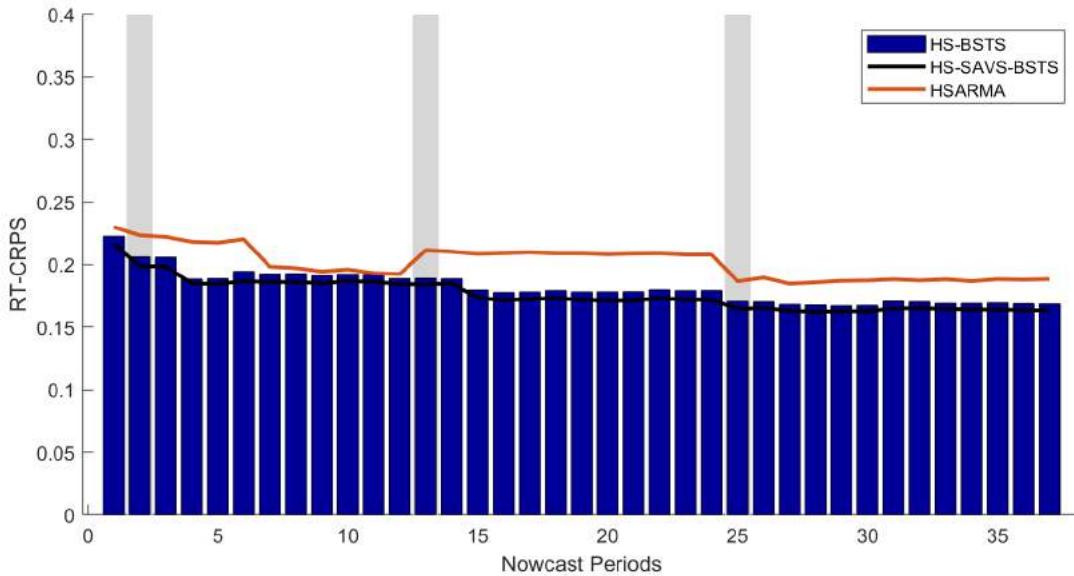


Figure 20: Real-Time-CRPS for 37 nowcast periods.

As can be clearly seen from Figures 18-20, the point as well as density nowcasts are clearly worse for the ARMA compared to the LLT horseshoe prior models.

This superior performance is related to the argument made above: the LLT model captures the slow-moving long-run growth component which gives the BSTS models the flexibility to capture deviations from this trend gathered from the explanatory information. This is underlined in figure (21): the LLT component captures smooth transitions in GDP growth, while the

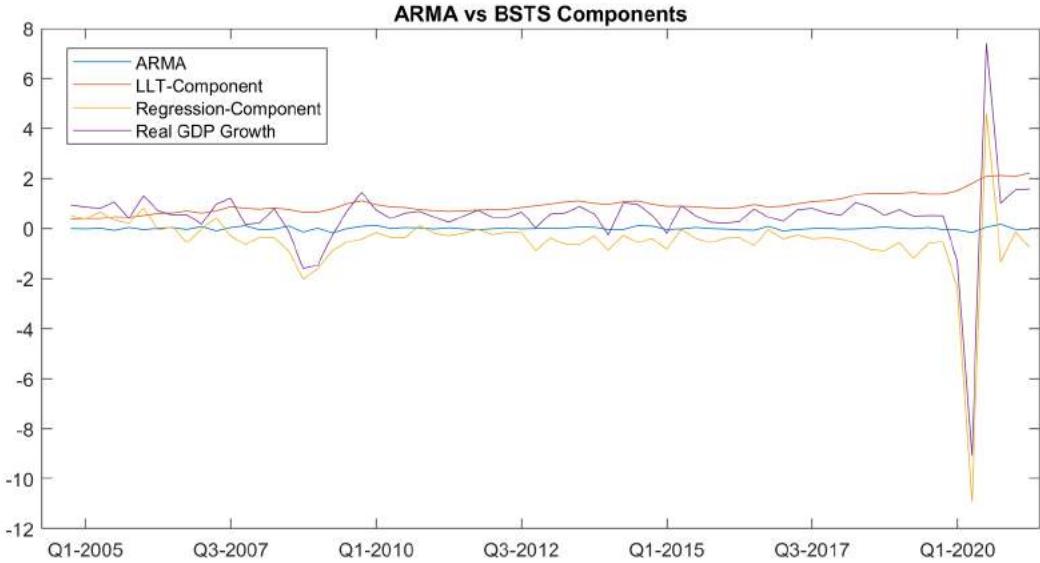


Figure 21: Trend and regression decomposition for the HS-BSTS model and the best performing HS-ARMA model (ARMA(1,1)).

large data set captures large troughs and peaks, such as during the financial crisis. By contrast, the ARMA component is of very small magnitude and displays erratic short-run movements. This suggests that short run dynamics are indeed better captured by the macro and Google Trend information.

A.4 State Space Estimation and Forecasting

A.4.1 Estimation

Assume analogously to A.1.1 and A.1.2 that all regression parameters have been sampled such that conditionally on β , we estimate $y - X\beta = \hat{y}_t = \tau_0 + \sigma_\tau \tilde{\tau}_t + t\alpha_0 + \sigma_\alpha \sum_{s=1}^t \tilde{\alpha}_s + \epsilon_t$, $\epsilon_t \sim N(0, \sigma_y)$ and $\tilde{\tau}_t = \tilde{\tau}_{t-1} + u_t^\tau$, $u_t^\tau \sim N(0, 1)$, $\tilde{\alpha}_t = \tilde{\alpha}_{t-1} + u_t^\alpha$, $u_t^\alpha \sim N(0, 1)$. Since the state processes $\{\tilde{\tau}, \tilde{\alpha}\}_{t=1}^T$ are independent of the other parameters in the non-centred formulation, we proceed by first estimating the states and then $\theta = \{\tilde{\tau}_0, \tilde{\alpha}, \sigma_\tau, \sigma_\alpha\}$.

States $\{\tilde{\tau}, \tilde{\alpha}\}_{t=1}^T$ can be sampled by any state space algorithm, e.g. Durbin and Koopman (2002), Carter and Kohn (1994) or Frühwirth-Schnatter (1994). We instead opt for the precision sampler by Chan (2017a) which exploits the joint distribution of the states which paired with sparse matrix operations yields significant increases in statistical as well as computational efficiency (Grant and Chan, 2017). Since $\tilde{\alpha}_s$ enters in the observation equation as a sum, we

define $\tilde{A}_t = \sum_{s=1}^t \tilde{\alpha}_s$. Notice that equation (3) implies that $\mathbf{H}\tilde{\alpha} = \tilde{\mathbf{u}}^\alpha$, where \mathbf{H} is the first difference matrix²⁸ and $\tilde{\mathbf{u}}^\alpha \sim N(\mathbf{0}, \mathbf{I}_T)$. Notice that $\tilde{A}_t = \tilde{\alpha}_1$ which implies that $\tilde{A}_t - \tilde{A}_{t-1} = \tilde{\alpha}_t$. Hence, this gives us back the desired $\mathbf{H}\tilde{\mathbf{A}} = \tilde{\alpha}$. Solving $\tilde{\mathbf{A}} = \mathbf{H}^{-1}\tilde{\mathbf{u}}^\alpha = \mathbf{H}^{-2}\tilde{\mathbf{u}}^\alpha$. Therefore

$$\tilde{\mathbf{A}} \sim N(\mathbf{0}, (\mathbf{H}^2' \mathbf{H}^2)^{-1}) \quad (30)$$

To sample the states jointly, define $\xi = (\tilde{\boldsymbol{\tau}}', \tilde{\mathbf{A}}')'$. Then \hat{y} can be re-written as:

$$\hat{y} = \tau_0 \mathbf{1}_T + \alpha_0 \mathbf{1}_{1:T} + X_\xi \xi + \epsilon, \quad (31)$$

where $\mathbf{1}_{1:T}$ is defined as $(1, 2, \dots, T)'$ and $\mathbf{X}_\xi = (\sigma_\tau I_T, \sigma_\alpha I_T)$. Since \mathbf{X}_ξ is a sparse matrix, manipulations in programs which utilise sparse matrix operations will be very fast.

Similar calculations result in the implicit prior $\tilde{\boldsymbol{\tau}} \sim N(\mathbf{0}, (\mathbf{H}' \mathbf{H})^{-1})$. Now, since by assumption $\boldsymbol{\tau}$ and $\tilde{\mathbf{A}}$ are independent, the combined for ξ is:

$$\xi \sim N(0, \mathbf{P}_\xi^{-1}), \quad (32)$$

where $\mathbf{P}_\xi = \text{diag}(\mathbf{H}' \mathbf{H}, \mathbf{H}^2' \mathbf{H}^2)$. The posterior is thus standard:

$$p(\xi | \mathbf{y}, \sigma_y^2) \sim N(\bar{\xi}, A_\xi^{-1}) \quad (33)$$

where $K_\xi = P_\xi + \frac{1}{\sigma_y^2} \mathbf{X}_\xi' \mathbf{X}_\xi$ and $\bar{\xi} = \mathbf{K}_\xi^{-1} (\frac{1}{\sigma_y^2} \mathbf{X}_\xi' (\mathbf{y} - \tau_0 \mathbf{1}_T - \alpha_0 \mathbf{1}_{1:T}))$.

Conditionally on ξ , θ are drawn by simple linear regression results, where we specify a generic prior covariance as $V_\theta = \text{diag}(1, 1, 0.1, 0.1)$ and prior mean $\theta_0 = (0, 0, 0, 0)$. In particular, define $X_\theta = (\mathbf{1}_T, \mathbf{1}_{1:T}, \tilde{\boldsymbol{\tau}}, \tilde{\mathbf{A}})$. Then:

$$\begin{aligned} \theta | \mathbf{y}, \sigma_y^2 &\sim N(\bar{\theta}, A_\theta^{-1}) \\ A_\theta &= (V_\theta^{-1} + \frac{1}{\sigma_y^2} X_\theta' X_\theta) \\ \bar{\theta} &= A_\theta^{-1} (V_\theta^{-1} \theta_0 + \frac{1}{\sigma_y^2} X_\theta' \hat{y}). \end{aligned} \quad (34)$$

A.4.2 Forecasting

Taking equation (1) as our starting point, it is well known that the predictive density $p(y_t | \mathbf{y}^{t-1}, \beta, \theta, \sigma_y^2)$, where $\mathbf{y}^{t-1} = (y_1, \dots, y_{t-1})$, can be generated by the Kalman filter. Since the state space is instead estimated by precision sampling, and hence, without Kalman recursions, the literature

²⁸[Put in here the first difference matrix as defines on p.81, Chan (2017)]

has proposed (1) conditionally optimal Kalman mixture approximations (Bitto and Frühwirth-Schnatter, 2019), (2) pure simulation based methods to approximate (1) (Belmonte et al., 2014), and (3) what Bitto and Frühwirth-Schnatter (2019) call naive Gaussian mixture approximation (see A.1.2.2 of Bitto and Frühwirth-Schnatter (2019)). In simulations as well as the empirical example we found that results are very similar independent of the sampling technique. For computationally simplicity we present here method (2).

The predictive on-step-ahead distribution $p(y_t | \mathbf{y}_{t-1})$ can be generated by first drawing from the non-centred states which with the draws of the other model parameters yield draws from the predictive. More specifically, for posterior draw $m = 1, \dots, M$:

1. Draw $(\tilde{\tau}_t^{(m)}, \tilde{\alpha}_t^{(m)})$ from $N(\tilde{\tau}_{t-1}^{(m)}, 1)$ and $N(\tilde{\alpha}_{t-1}^{(m)}, 1)$ respectively
2. Generate $\alpha_t = \alpha_0^{(m)} + \sigma_\alpha^{(m)} \tilde{\alpha}_t^{(m)}$ and $\tau^{(m)} = \tau_0^{(m)} + \sigma_\tau^{(m)} \tilde{\tau}_t^{(m)} + t \alpha_0^{(m)} + \sigma_\alpha^{(m)} \sum_{s=1}^t \alpha_s^{(m)}$
3. Generate $x'_t \beta^{(m)} + \tau^{(m)} + \sigma_y^{(m)} u$, where $u \sim N(0, 1)$

To obtain an approximation to the continuous approximation to $p(y_t | \mathbf{y}^{t-1})$, one can then use a kernel density smoother such as "kdensity" in Matlab.

A.5 BSTS-t Model

The model to be estimated is:

$$y_t = \tau_0 + \sigma_\tau \tilde{\tau}_t + t \alpha_0 + \sigma_\alpha \sum_{s=1}^t \tilde{\alpha}_s + x'_t \beta + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma^2 \psi_t), \quad \psi_t \sim \mathcal{G}^{-1}(\eta/2, \eta/2) \quad (35)$$

and

$$\begin{aligned} \tilde{\tau}_t &= \tilde{\tau}_{t-1} + \tilde{u}_t^\tau, \quad \tilde{u}_t^\tau \sim N(0, 1) \\ \tilde{\alpha}_t &= \tilde{\alpha}_{t-1} + \tilde{u}_t^\alpha, \quad \tilde{u}_t^\alpha \sim N(0, 1) \end{aligned} \quad (36)$$

Compared to the normal BSTS models,, $\boldsymbol{\psi} = (\psi_1, \dots, \psi_T)'$ and η are additional unknown parameters.

For exposition, we treat η for now as known. The Gibbs sampler to draw inference on this model needs the conditional posteriors: $p(\beta | y, \theta, \tilde{\tau}, \tilde{\alpha}, \sigma_y^2, \boldsymbol{\psi})$, $p(\tilde{\tau}, \tilde{\alpha} | y, \beta, \theta, \sigma_y^2, \boldsymbol{\psi})$, $p(\theta | y, \tilde{\tau}, \tilde{\alpha}, \sigma_y^2, \boldsymbol{\psi})$ and $p(\sigma_y^2 | y, \theta, \tilde{\tau}, \tilde{\alpha}, \boldsymbol{\psi})$.

To derive $p(\beta | \bullet)$, rewrite again $y^* = y - \tau = X\beta + \epsilon$. Then $\epsilon_t \sim (0_T, \sigma_y^2 \Theta)$, where $\Theta = \text{diag}(\psi_1, \dots, \psi_T)$. Defining $\Sigma = \sigma^2 \Theta$:

$$\begin{aligned}
\beta | \bullet &\sim N(\hat{\beta}_1, D_1) \\
D_1 &= (\Lambda_*^{-1} + \frac{1}{\sigma_y^2} X' \Sigma^{-1} X)^{-1} \\
\hat{\beta}_1 &= A^{-1} \frac{1}{\sigma_y^2} \Sigma^{-1} y^*,
\end{aligned} \tag{37}$$

where the diagonal of Λ_* is populated by the shrinkage scales of either the horseshoe prior or the SSVS.²⁹

To sample from $p(\theta | \bullet)$, rewrite again $\hat{y} = y_t - x_t' \beta = \tau_0 + \sigma_\tau \tilde{\tau}_t + t \alpha_0 + \sigma_\alpha \sum_{s=1}^t \tilde{\alpha}_s + \epsilon_t$, $\epsilon_t \sim N(0, \sigma_y^2 \psi_t)$ and $\psi_t \sim \mathcal{G}^{-1}(\eta/2, \eta/2)$. Again, define $X_\theta = (\mathbf{1}_T, \mathbf{1}_{1:T}, \tilde{\tau}, \tilde{\mathbf{A}})$. Then the posterior is:

$$\begin{aligned}
\theta | \bullet &\sim N(\hat{\beta}_2, D_2) \\
D_2 &= (V_\theta^{-1} + \frac{1}{\sigma_y^2} X_\theta' \Sigma^{-1} X_\theta)^{-1} \\
\hat{\beta}_2 &= D_2 (V_\theta^{-1} \theta_0 + \frac{1}{\sigma_y^2} X_\theta' \Sigma^2 \hat{y})
\end{aligned} \tag{38}$$

To sample from $(\tilde{\tau}, \tilde{\alpha} | \bullet)$, take equation 31 with $\epsilon_t \sim N(0, \sigma_y^2 \psi_t)$ and $\psi_t \sim \mathcal{G}^{-1}(\eta/2, \eta/2)$. Then the posterior with the same steps as above is:

$$\begin{aligned}
\xi | \hat{y}, \sigma_y^2 &\sim N(\hat{\beta}_3, D_3) \\
D_3 &= (P_\xi + \frac{1}{\sigma_y^2} X_\xi' \Sigma^{-1} X_\xi)^{-1} \\
\hat{\beta}_3 &= \mathbf{K}_\xi^{-1} \left(\frac{1}{\sigma_y^2} X_\xi' \Sigma^{-1} (\mathbf{y} - \tau_0 \mathbf{1}_T - \alpha_0 \mathbf{1}_{1:T}) \right).
\end{aligned} \tag{39}$$

To derive the posterior $p(\psi | \bullet)$, notice that each ψ_t is univariate and independently distributed. Hence:

$$p(\psi | \bullet) \propto \prod_{t=1}^T \psi_t^{-\frac{\eta+1}{2} + 1} e^{-\frac{1}{2\psi_t} \frac{(\eta+y_t - \tau - x_t' \beta)^2}{\sigma_y^2}}. \tag{40}$$

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

$$\psi_t \sim \mathcal{G}^{-1}\left(\frac{\eta+1}{2}, \frac{1}{2} \frac{(y_t - \tau - x_t' \beta)^2}{\sigma_y^2}\right) \tag{41}$$

²⁹Since the horseshoe uses a conjugate formulation, the posterior moments for the horseshoe are $D_1 = \sigma^2(X' \Sigma^{-1} X + \Lambda^{-1})$ and $\hat{\beta}_1 = D_1^{-1} X' \Sigma^{-1} y^*$.

Finally, regarding the unknown η , we specify a uniform prior $\eta \sim U[2, 50]$. 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:

$$\begin{aligned} p(\eta | \bullet) &\propto p(\psi p(\eta)) \\ &\propto \prod_{t=1}^T \frac{(\eta/2)^{\frac{\eta}{2}}}{\Gamma(\eta/2)} \psi_t^{-(\frac{\eta}{2}+1)} e^{-\frac{\eta}{2\psi_t}} \\ &= \frac{(\eta/2)^{\frac{T}{2}}}{\Gamma(\eta/2)^T} \left(\prod_{t=1}^T \right)^{-(\frac{\eta}{2}+1)} e^{\frac{\eta}{2} \sum_{t=1}^T \psi_t^{-1}} \end{aligned} \quad (42)$$

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 Metropolis-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

$$\alpha(X, Y) = \min \left\{ \frac{f(Y)g(X)}{f(X)g(Y)}, 1 \right\}, \quad (43)$$

where $g(\cdot)$ is the proposal density. In order for the Metropolis-Hastings sampler to quickly explore the typical set of $\eta | \bullet$, 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).

The sampling algorithm follows the same sequence as in main text of the paper, however with steps $p(\psi | \bullet)$ and $p(\eta | \bullet)$ added before sampling $\sigma_y^2 | \bullet$ is sampled.

B Graphs

B.1 HS-BSTS-t Model Results

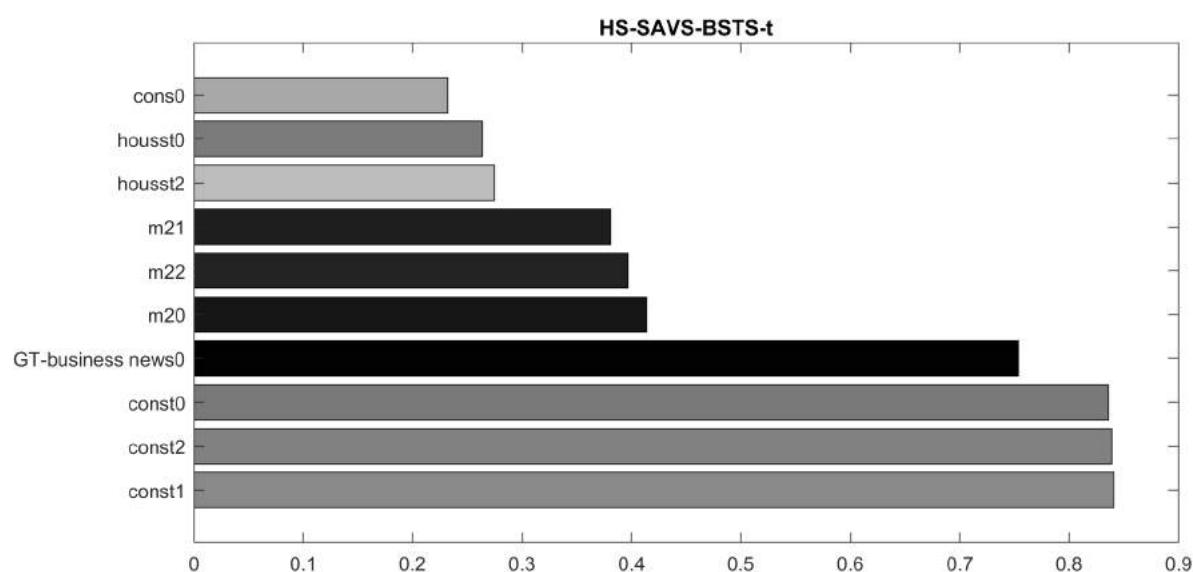


Figure 22: Posterior inclusion probabilities for the HS-SAVS-BSTS-t model.

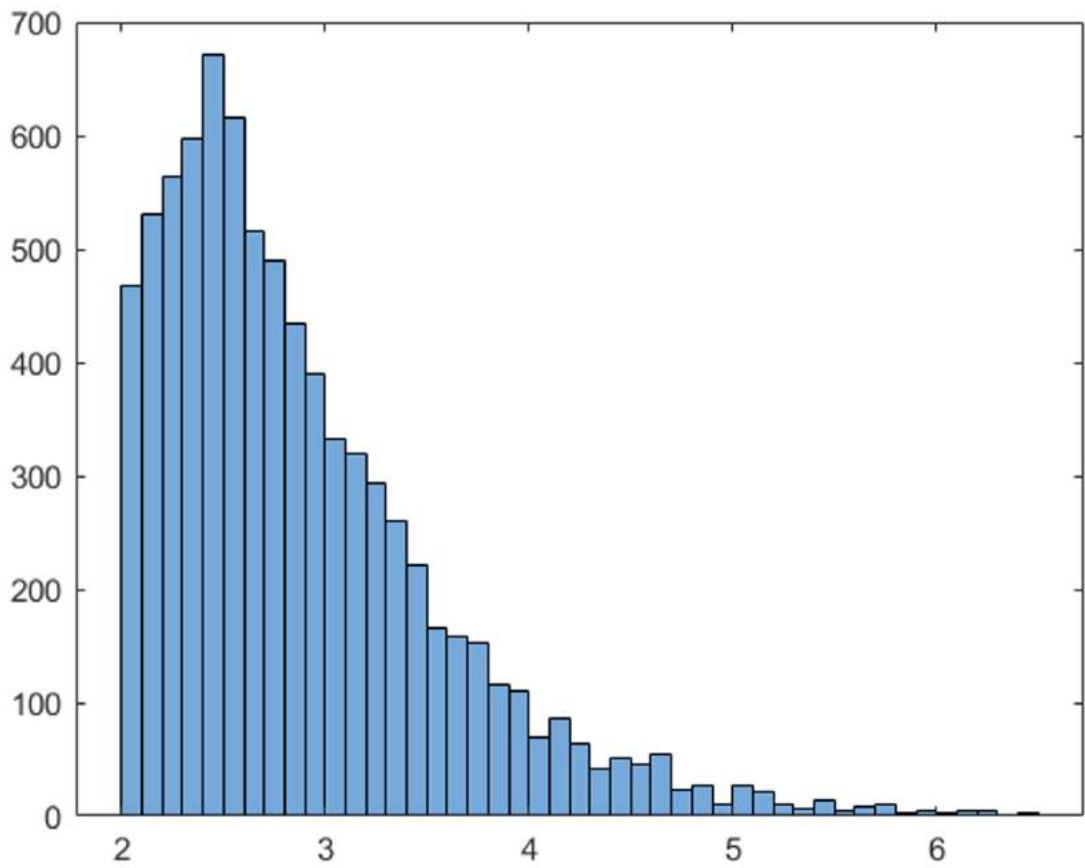


Figure 23: Posterior of the degree of freedom parameter, η .

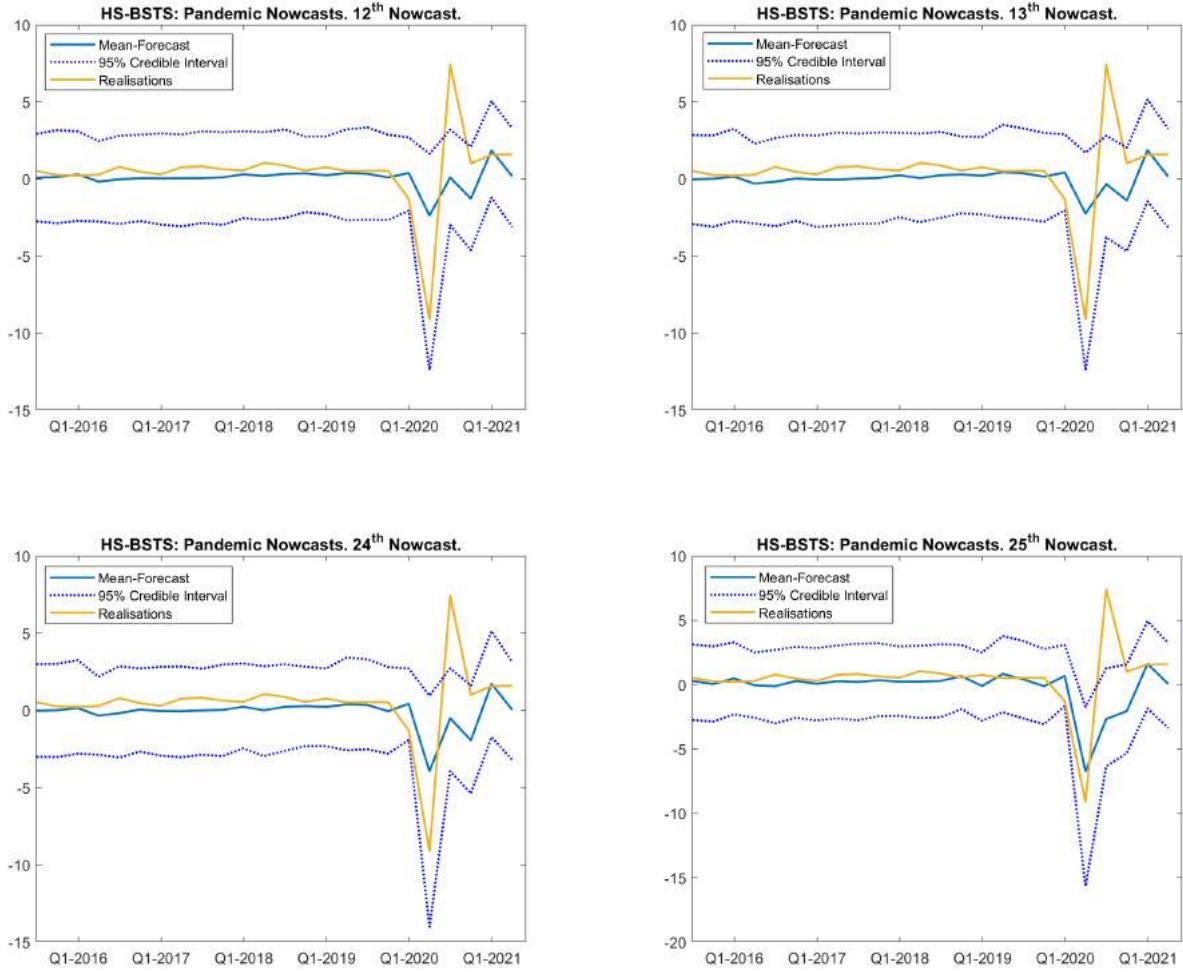


Figure 24: Predictive distributions for full HS-BSTS-t model for nowcast periods 12,13,24 and 25. Periods 12 and 24 are prior to Google Trends releases.

B.2 In-Sample Results

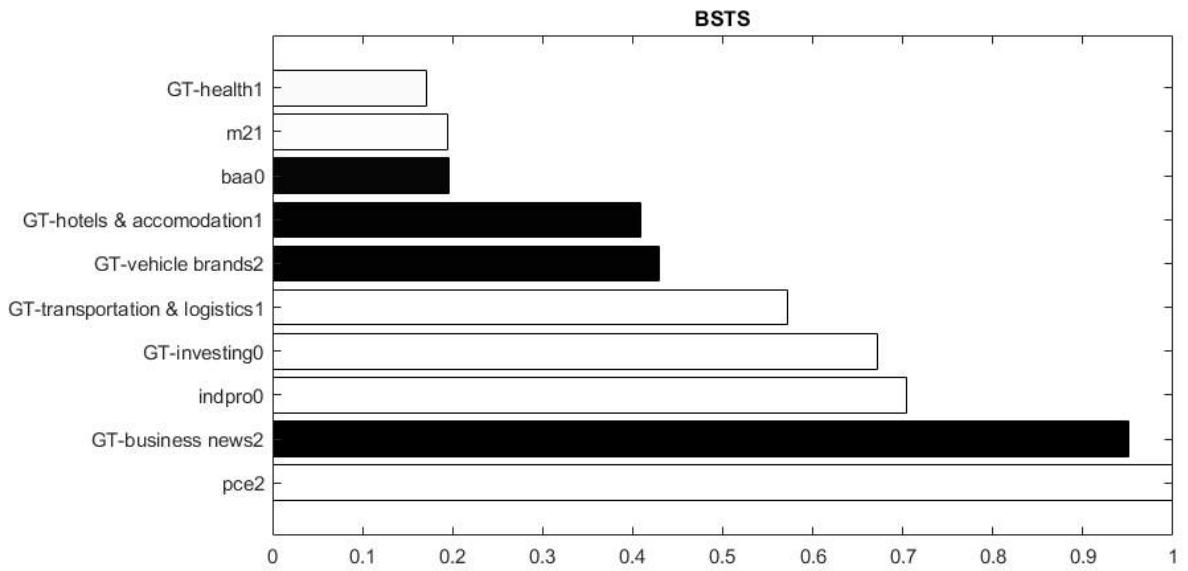


Figure 25: Posterior inclusion probabilities for the original BSTS model of Scott and Varian (2014).

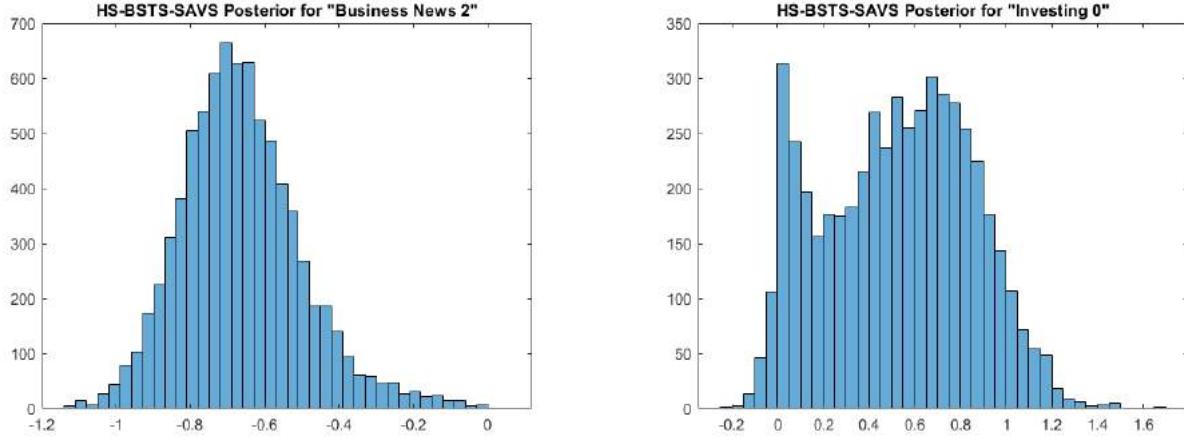


Figure 26: Posterior of the Google category 'Business news 2' (left) and topic 'Investing 0' for HS-SAVS-BSTS.

B.3 Google Topic/Category Plots

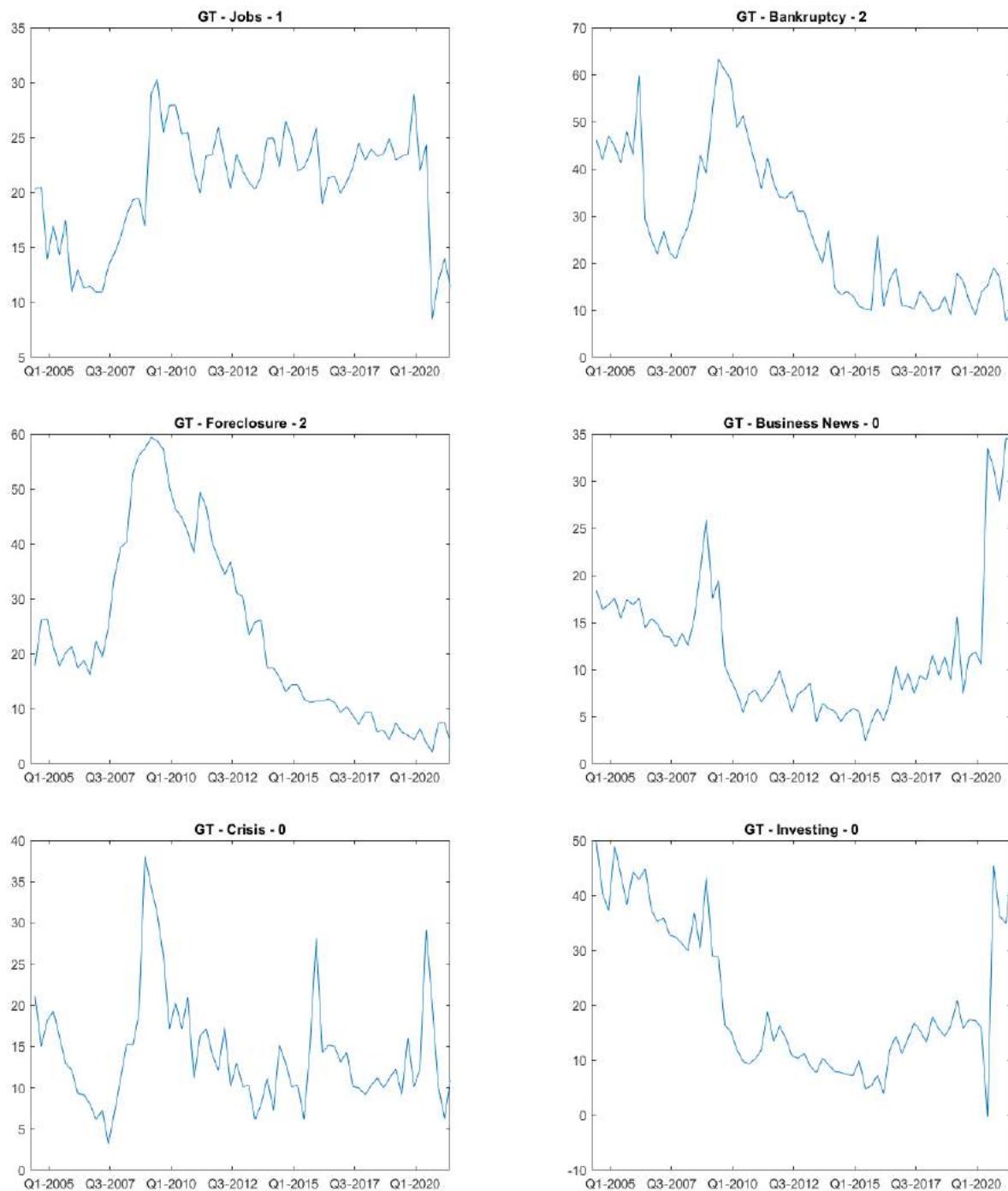


Figure 27: U-MIDAS transformed Google search data plots.