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

Galina Potjagailo[1], David Kohns[2]

[1]Bank of England
[2]Aalto University, Heriot-Watt University, European Central Bank, Centre for Energy Economics Research and Policy (CEERP)

EEA-ESEM Congress 2023
Barcelona School of Economics
August 31, 2023

Motivation & What we do

- Nowcast and forecast models face challenges in presence of large and heterogeneous shocks and non-linearities or shifts over time
  - Large observations can distort parameter estimates and increase uncertainty
  - Changing correlation structure and heterogeneous dynamics across indicators

- Require flexible model features
  - Accounting for time-varying trends and stochastic volatilities (SV) beneficial in UC, DFM, VAR models (Stock and Watson, 2009; Clark, 2011; Antolin-Diaz et al., 2017)
  - Covid-19: Account for extreme observations via t-distr. errors or outliers (Carriero et al., 2021; Lenza and Primiceri, 2022; Antolin-Diaz et al., 2021)
  - Bayesian shrinkage (Carriero et al., 2015; Mogliani and Simoni, 2021)

- We combine time-varying components with multivariate MIDAS (Ghysels et al., 2020) and Bayesian group-shrinkage: Trend-SVt-BMIDAS with GIGG prior

Literature

Trend-SV-t-BMIDAS model

four key features

1. Time-varying unobserved components in the lower-frequency target variable (time-varying Trend, stochastic volatlity (SV), t-distr. errors)
2. Information from high frequency indicators in multivariate MIDAS regression
3. Bayesian group-shrinkage via a global-local prior with three tiers of continuous shrinkage (overall, between indicators, and within lags of an indicator)
4. A new group-wise sparsification algorithm on the posterior
    - Ex-post sparsification motivated by decision theory allows for variable selection and helps interpret signals over time via inclusion probabilities.
    - Approach separates shrinkage and sparsity akin to "illusion of sparsity" (Giannone et al., 2021), while accounting for within-group correlation.

Main Results: empirical application for nowcasting UK GDP growth

- Combination of time-varying components and group-shrinkage GIGG prior improves nowcast performance for UK GDP growth, before and including the pandemic.

  1. GIGG group-shrinkage prior shrinks information set towards a sparse selection indicators, while to a lesser extent also drawing on other indicators.
  2. Time-varying components help to shift between the most meaningful indicators over data release cycle, rather then relying on constant signals.

- Inclusion probabilities inform about signals exploited by the model

  - Early in data release cycle reliance on surveys, then shift to 'hard' indicators.
  - Covid-19 pandemic: shift towards indicators for services, away from production.
  - Other shrinkage priors such as Horseshoe rely on diffuse or broad set of indicators and profit less from time-varying components.

Methodology

The Trend-BMIDAS-SV-t Model

$$
\begin{aligned}
y_t &= \tau_t + \sum_{k=1}^{K} \mathcal{B}(L^{1/m}; \theta_k) x_{k,t} + \sqrt{\lambda_t} e^{\frac{1}{2}(h_0 + w_h \tilde{h}_t)} \tilde{\epsilon}_t^y, \\
\tilde{\epsilon}_t^y &\sim N(0,1), \ \lambda_t \sim IG(\nu/2, \nu/2) \\
\tau_t &= \tau_{t-1} + e^{\frac{1}{2}(g_0 + w_g \tilde{g}_t)} \tilde{\epsilon}_t^\tau, \ \tilde{\epsilon}_t^\tau \sim N(0,1) \\
\tilde{h}_t &= \tilde{h}_{t-1} + \tilde{\epsilon}_t^h, \ \tilde{\epsilon}_t^h \sim N(0,1) \\
\tilde{g}_t &= \tilde{g}_{t-1} + \tilde{\epsilon}_t^g, \ \tilde{\epsilon}_t^g \sim N(0,1).
\end{aligned}
\tag{1}
$$

- τ_t: time-varying trend; τ_t and y_t are lower frequency (quarterly)
- $B(L^{1/m}, \theta_k) x_{t,k} = \sum_{j=1}^{m} \omega(\frac{j-1}{m}; \theta_k) x_{t-(j-1)/m,k}$
  - θ: $(p_k + 1) * K$ parameters that link higher and lower frequency observations.
  - $\omega : \mathbb{R} \times \mathbb{R}^L \to \mathbb{R}$, nests Almon and U-MIDAS Almon
- h_t, g_t: SVs, non-centered (Frühwirth-Schnatter and Wagner, 2010)
- λ_t: enforces a ν-degrees of freedom t-distribution, fat-tailed SV

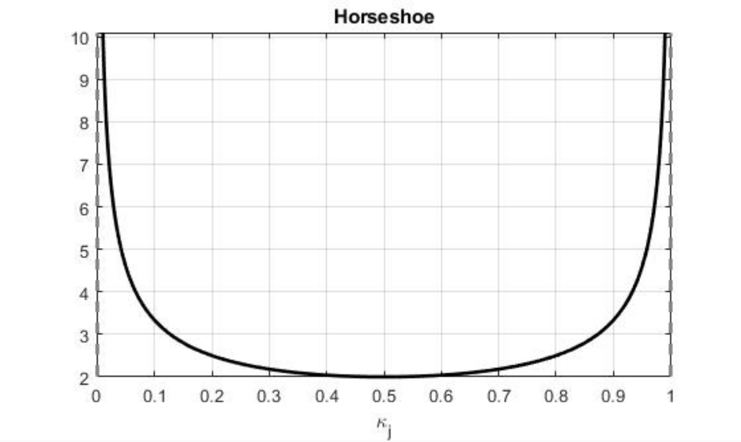
Group-shrinkage prior on multivariate MIDAS component

- GIGG (Group-Inverse-Gamma-Gamma) global-local prior (Boss et al., 2021)
  - Models group-shrinkage + correlation among higher frequency lags
- Each group g has $p_g + 1$ parameters to estimate

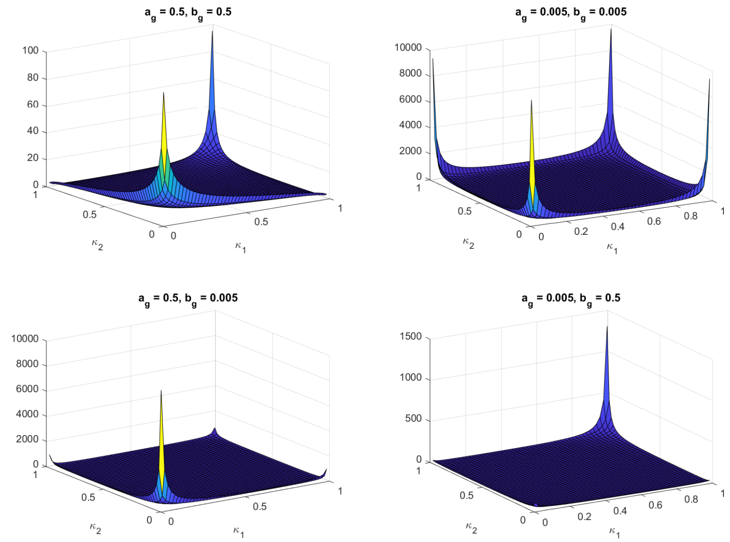
$$\theta_{g,i} \sim N(0, \vartheta^2 \gamma_g^2 \varphi_{g,i}^2), \ \ \forall i \in \{0, \cdots, p_g + 1\}$$
$$\vartheta \sim C_+(0,1), \ \ \gamma_g^2 | a_g \sim G(a_g, 1), \ \ \varphi_{g,i}^2 | b_g \sim IG(b_g, 1), \tag{2}$$

  - ϑ controls the overall level of sparsity
  - $\gamma_g^2(a_g)$ controls sparsity across groups g
  - $\varphi_{g,i}^2(b_g)$ controls correlation of group members i within g
- We set $g = k$, i.e. groups defined as lags of each indicator (can be extended to groupings across indicators if k large)
- At group-size 1, $a_g = b_g = 0.5$, give horseshoe prior (Carvalho et al., 2010)

Univariate shrinkage with global-local prior



GIGG Prior Visualisation: Bi-variate shrinkage hyperparamers



Group-sparsification step on the posterior

- Lack of interpretability with continuous priors: posterior coefficients remain non-zero (Hahn and Carvalho, 2015), impact from indicators on nowcast opaque.
- Solution: Ex-post sparsification algorithm to the posterior θ_g.
  - Decision tool that is separate from regularisation imposed by prior.
  - Minimise a utility function over Euclidean distance between a linear model which penalises group-size (akin to Zou, 2006) and the model's prediction:

$$\mathcal{L}(\tilde{Y}, \alpha) = \frac{1}{2}||XW\alpha - \tilde{Y}||_2^2 + \sum_{k=1}^{K} \phi_k ||\alpha_k||_2, \tag{3}$$

  - penalisation term creates a soft-thresholding effect between [-ϕ_k, ϕ_k]
  - finds smallest subset of groups to achieve predictive performance closest to unsparsified model, coefficients in other groups forced to zero.
- Gives inclusion probabilities that inform about the relative impact of an indicator
  - relative frequency of lag-group k in the sparsified estimate $\alpha^{*(s)}$ over Gibbs draws

Other priors and estimation

Priors for latent states standard: $(\boldsymbol{\tau}, \tilde{\boldsymbol{h}}, \tilde{\boldsymbol{g}})$ joint normal prior as in Chan and Jeliazkov (2009), McCausland et al. (2011) and Kim et al. (1998)

Estimation via Metropolis-within-Gibbs sampler M-H Gibbs

- Recursive sampling from conditional distributions: MIDAS parameters θ, GIGG hyperparameters, latent states $(\boldsymbol{\tau}, \tilde{\boldsymbol{h}}, \tilde{\boldsymbol{g}})$ (non-recursively, as in Chan and Jeliazkov (2009)), λ_t, degrees of freedom ν
- sampling of ν requires Metropolis step
- 5000 burn-in iterations, retain further 5000 for inference

Empirical Application

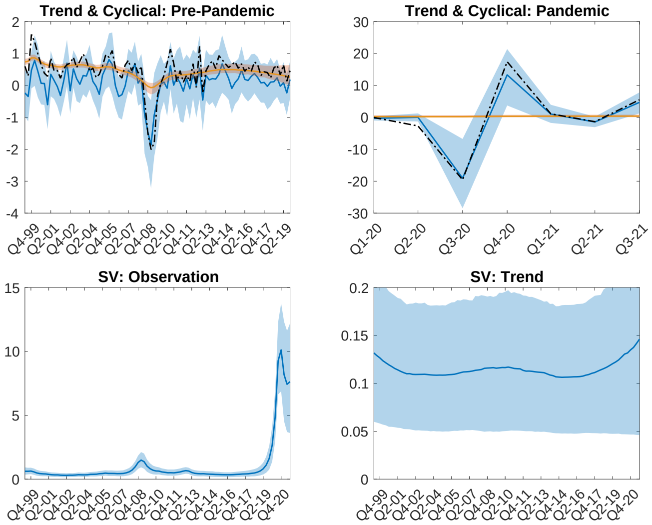
Nowcast quarterly UK GDP growth, 1999-2021

- Setup
  - In-Sample Start: Q1 1999, Nowcast Start: Q1 2011
  - Nowcast End: "pre-pandemic" Q4 2019, "including pandemic" Q3 2021
- Monthly indicators
  - indices of services and production, trade
  - surveys (CBI, PMI, GfK)
  - labour market (unemployment rate, employment, vacancies, hours)
  - mortgage approvals, VISA consumer spending
- Nowcast evaluation
  - pseudo-real-time calendar: 20 nowcasts per quarter around data releases calendar
  - each nowcast has new information set, latest available 6 monthly obs. of each indicator
- Metrics
  - Point: Root-mean-squared forecast error (RMSFE)
  - Density: Average cumulative rank probability score (CRPS)

Three Sets of Results

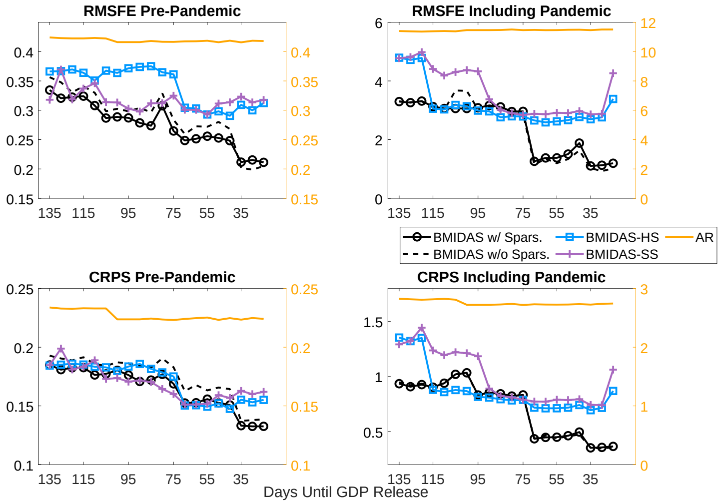
1. Posterior estimates time-varying trend, cyclical component, and stochastic volatilities
2. Nowcast evaluation
   - prior alternatives, all with time varying components
   - prior alternatives, without time varying components
3. Inclusion probabilities Signals exploited over the data release cycle and over time

Time-varying components - Posterior estimates

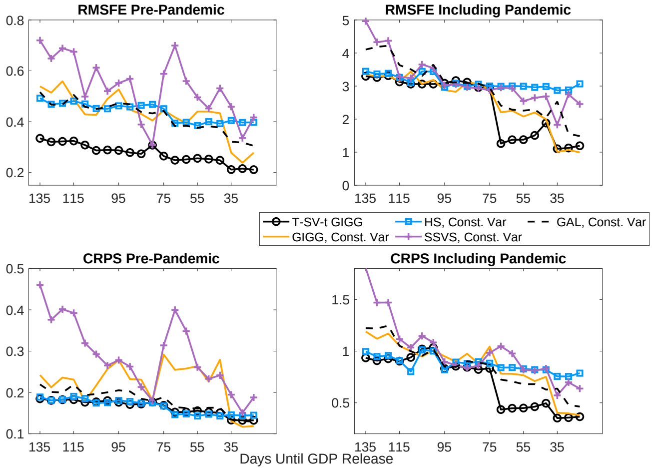


sensitivity

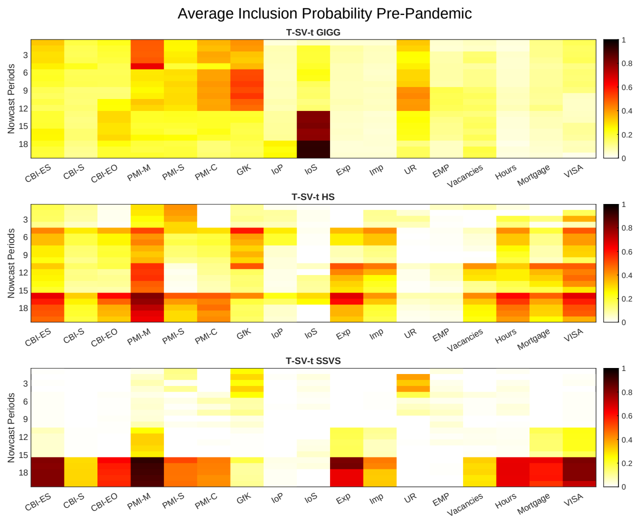
Role of group-shrinkage - Nowcast evaluation across prior choices



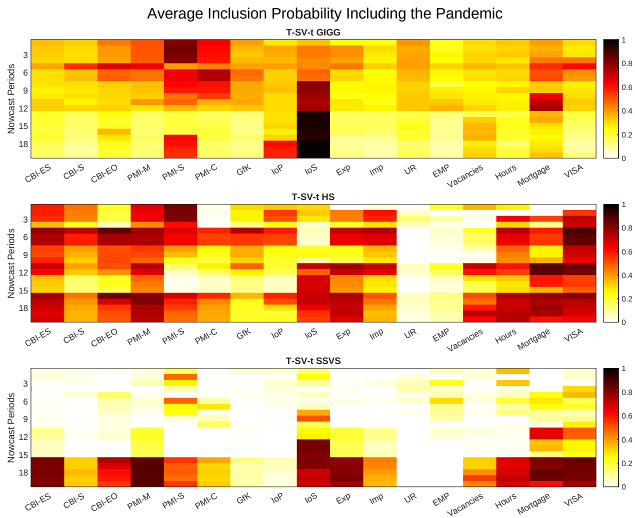
Role of group-shrinkage prior + shutting down time-variation



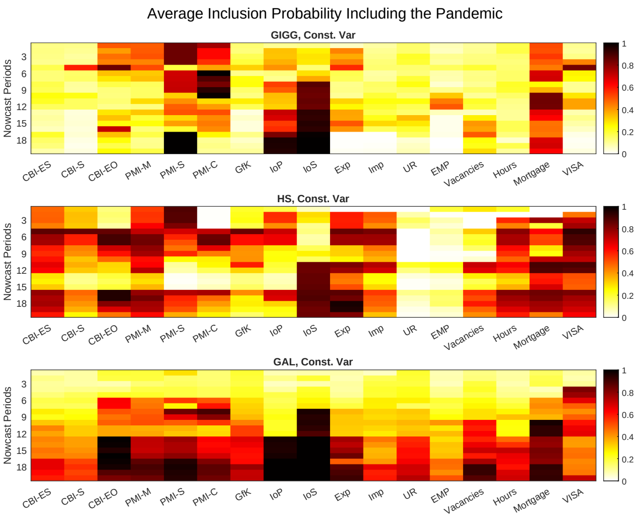
Intuition - signal extraction across priors (pre-pandemic period)



Intuition - signal extraction across priors (incl. pandemic period)



Intuition - signal extraction, no time-varying components (incl. pandemic period)



Conclusions

- Time-varying components + Bayesian MIDAS + flexible group-shrinkage prior.
    - Group-shrinkage and time-varying components (trends, volatilities, large errors) jointly lead to strong nowcasting performance for UK GDP growth.
- Approach brings new insights to literature on density vs sparsity in macroeconomic forecasting (Giannone et al., 2021)
    - grouping + time variation in shrinkage important to separate the relevant sub-group of indicators in each nowcast period and over time (e.g. Covid-19)
    - Shrinkage and sparsification are split apart: no a priori sparsification via prior, but on the posterior to enhance interpretability
- Can be relevant to a range of exercises where group shrinkage matters (e.g. also disaggregated data).

Thank you

Contact: galina.potjagailo@bankofengland.co.uk

Appendix

Links to the literature

- The model nests / compares to existing models when shutting down model features
  - multivariate BMIDAS with horseshoe (Kohns and Bhattacharjee, 2022) or spike-and-slab group shrinkage prior (Mogliani and Simoni, 2021) asymptotically
  - BMIDAS model with SV (Carriero et al., 2015)
  - Trend-SV-outl. DFM (Antolin-Diaz et al., 2021)

- MIDAS literature (Ghysels et al., 2007, 2020; Foroni et al., 2015)
- Machine learning for nowcasting (Babii et al., 2022)
- Global-local shrinkage priors (Polson and Scott, 2010; Polson et al., 2014; Carvalho et al., 2010) and group-lasso priors (Casella et al., 2010; Xu and Ghosh, 2015) and spike-and-slab (Ishwaran et al., 2005; Piironen et al., 2017)

back

Almon lag polynomial restricted MIDAS

- U-MIDAS (Foroni et al., 2015) involves many parameters and can lead to erratic weight profiles
- Restrict coefficients via Almon lag-polynomials on θ_i: assuming a $p_k << L_k$ polynomial process of the coefficients across high-frequency observations

Almon Lag MIDAS

Assume lags $i = 0, \cdots, L$ can be represented by a 3rd degree polynomial, then each HF parameter process, θ_i can be written as:

$$\theta_i = \beta_0 + \beta_1 i + \beta_2 i^2 + \beta_3 i^3 \tag{4}$$

We add economically relevant restrictions (Smith and Giles, 1976)

$$\begin{aligned} \theta'_L &= 0 \\ \theta_L &= 0 \end{aligned} \tag{5}$$

But: Smoothness of Almon-polynomial increases parameter correlation back

Metropolis-within-Gibbs sampling algorithm

1. Sample $\theta|\bullet \sim p(\theta|\mathbf{y}, \bullet)$
2. Sample hyper-parameters $\vartheta, \gamma_k^2, \varphi_{kj}^2, \nu_p$ in one block
   2.1 $\vartheta^2 \sim p(\vartheta^2|\mathbf{y}, \bullet)$
   2.2 $\gamma_k^2 \sim 1/p(\gamma_k^{-2}|\mathbf{y}, \bullet)$
   2.3 $\varphi_{kj}^2 \sim p(\varphi_{kj}^2|\mathbf{y}, \bullet)$
3. sample $\tilde{\boldsymbol{\tau}} \sim p(\tilde{\boldsymbol{\tau}}|\mathbf{y}, \bullet)$ and $\tau_0 \sim p(\tau_0|\mathbf{y}, \bullet)$
4. sample $\tilde{\mathbf{h}} \sim p(\tilde{\mathbf{h}}|\mathbf{y}, \bullet)$, $h_0 \sim p(h_0|\mathbf{y}, \bullet)$ and $w_h \sim p(w_h|\mathbf{y}, \bullet)$
5. sample $\tilde{\mathbf{g}} \sim p(\tilde{\mathbf{g}}|\mathbf{y}, \bullet)$, $g_0 \sim p(h_0|\mathbf{y}, \bullet)$ and $\sim p(w_g|\mathbf{y}, \bullet)$
6. Sample $\{\lambda_t\}_{t=1}^T \sim p(\lambda_t|\mathbf{y}, \bullet)$
7. Sample $\nu \sim p(\nu|\mathbf{y}, \bullet)$ with a Metropolis step back

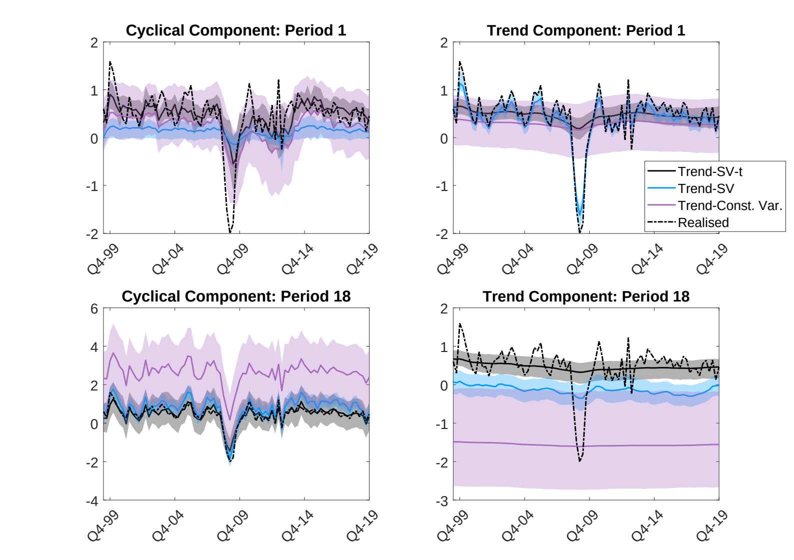
- sampling technique of Chan and Jeliazkov (2009) allows drawing steps 3.-5. in a non-recursive fashion which increases efficiency and can be sped up using sparse-matrices

Pseudo Real Time Calendar for UK Nowcast Application

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

back

Trend and SV posterior estimates, alternative models.



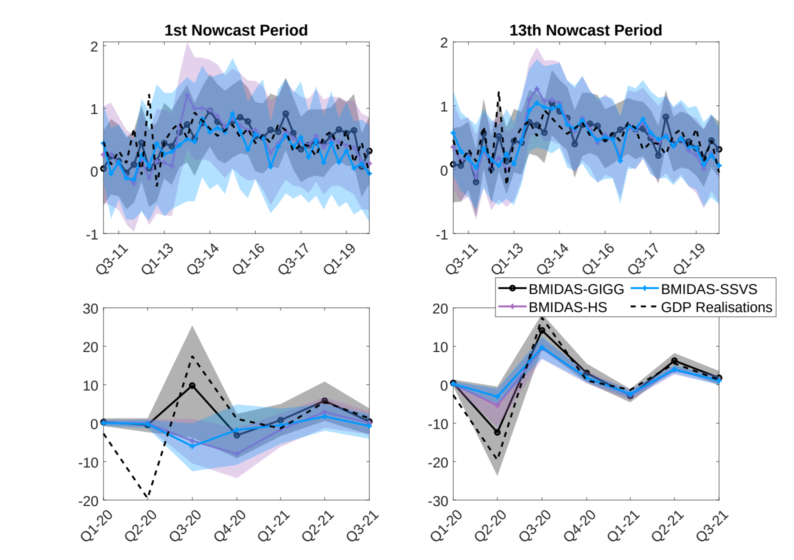
back

Point nowcast evaluation results, RMSFE relative to AR(2).

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

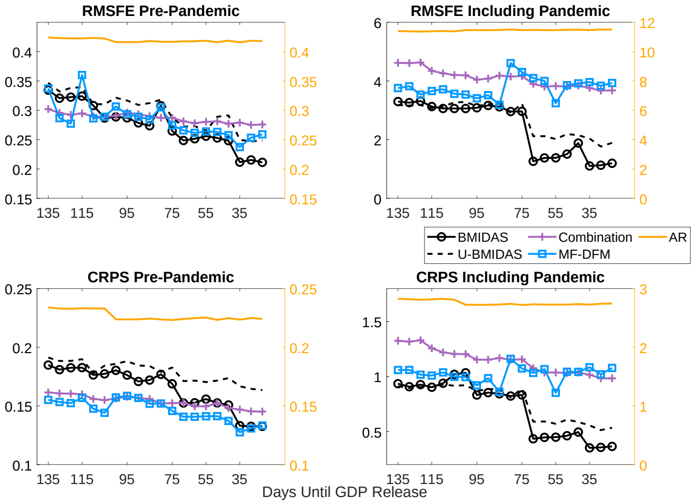
back

Posterior mean and density nowcasts, alternative priors.



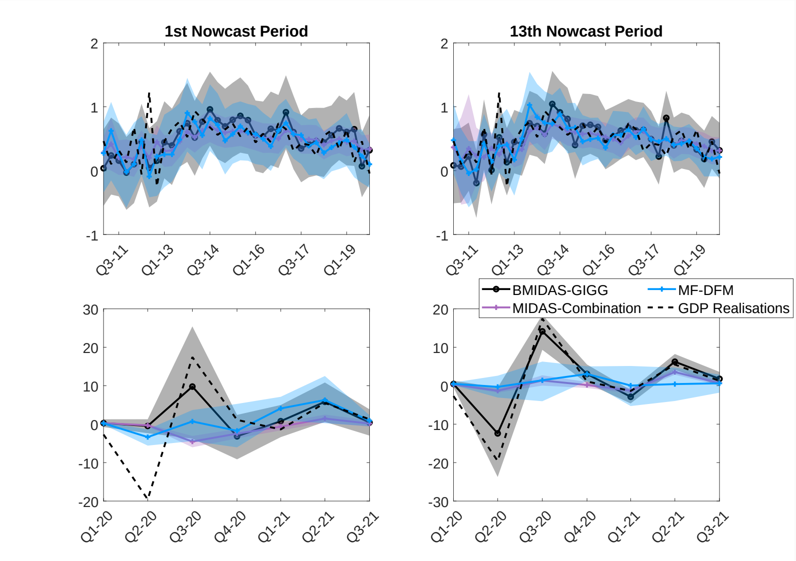
back

2b) Performance against alternatives to multivariate MIDAS



back

Posterior mean and density nowcasts, alternatives to BMIDAS.



back

References

References i

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

McCausland, W. J., Miller, S., and Pelletier, D. (2011). Simulation smoothing for state–space models: A computational efficiency analysis. Computational Statistics & Data Analysis, 55(1):199–212.

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

Piironen, J., Vehtari, A., 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 Scott, J. G. (2010). Shrink globally, act locally: Sparse Bayesian regularization and prediction. Bayesian Statistics, 9:501–538.

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

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

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

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

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