

High-Dimensional Bayesian Methods for Interpretable Nowcasting and Risk Estimation



David Eric Kohns

School of Social Sciences

Heriot-Watt University

A thesis submitted for the degree of

Doctor of Philosophy

01-2023

The copyright in this thesis is owned by the author. Any quotation from the thesis or use of the information contained in it must acknowledge this thesis as a source of the quotation or information.

To my parents and partner

Declaration

I, David Kohns, confirm that the work presented in this thesis is my own, except where explicitly stated otherwise.

Chapter 2 of my thesis is co-authored with Prof. Arnab Bhattacharjee (Heriot-Watt University & NIESR). The idea and topic emerged in much simpler form from my MSc dissertation, and was proposed in its current form by myself. I collected the data, developed the methodology and code, designed and ran the simulations, conducted the empirical analysis and wrote the paper as well as responded to reviewers during the publication process. Prof. Bhattacharjee was instrumental in guiding the writing process as well as editing both the final version of the paper as well the responses to reviewers. An earlier version of this paper was published on ‘arXiv’ under the title “Developments on the Bayesian Structural Time Series Model: Trending Growth”.

Chapter 3 of my thesis is co-authored with Dr. Galina Potjagailo (Bank of England). I developed the idea for the project as part of my PhD internship proposal to the Bank of England. The raw data for the empirical application was provided by the Bank of England via their FAME data warehousing software. I was responsible for developing methodology and writing the code, curating the data (based on similar code Dr. Potjagailo provided), conducting the empirical analyses, generating figures and tables, designing and implementing the simulation exercise, and writing the paper. Dr. Potjagailo contributed tremendously by advising on exposition of methodology, editing the writing, providing insights into specificities of the UK related data and helping curate the economic intuition behind the results.

Chapters 4 and 5 of my thesis are co-authored with PhD candidate Tibor Szendrei (Heriot-Watt University & NIESR). I brought into these papers knowledge on Bayesian estimation, shrinkage priors and variable selection, while Tibor brought in knowledge of quantile regression. Methodology and simulations were developed in tandem along our expertise, model estimation code was written by myself, along with code for conducting the empirical analyses. Tibor wrote code for the simulations. Data was downloaded and prepared via existing code, incorporated by Tibor. Both he and I were responsible for writing, editing and responding to reviewers. Earlier versions of these chapters have been published on ‘arXiv’ under the titles “Horseshoe Prior Bayesian Quantile Regression”, “Decoupling Shrinkage and Selection for the Bayesian Quantile Regression”.

Acknowledgements

First and foremost I would like to thank my supervisors Arnab Bhattacharjee, Mark Schaffer and Erkal Ersoy for their continuous support over the course of my PhD. Arnab somehow always found time for calls, long and short, even during the pandemic and his many travels. I also thank Mark for his prudent advice on many topics, not always related to research, which is more than one can hope for from an academic supervisor. Thank you also to Erkal for letting me participate in the BP energy project where I learned to look over the edge of my research plate (German proverb).

I thank my fellow PhD colleagues, Chris Aitken and Tibi Szendrei for not only being good friends and co-authors, but also listening to my rants. Our discussions and walks kept me sane during the PhD. A special thanks also to Tibi who was an invaluable co-author on many research projects, who encouraged and pushed our work on-wards with unrelenting positivity.

Many thanks also to Galina Potjagailo for all the invaluable and inspiring cooperation and guidance, from my research visit at the Bank of England, to the joint work which became an integral part of this PhD thesis. Her tenacity and attention to detail has motivated me to become a better researcher.

Without my parents and my partner, I would also not have made it to this stage of my PhD. They listened to my incessant self-doubt and provided all the right words of encouragement which enabled me to carry on. I have to especially thank my partner Ann for making me realise that there is more to life than research.

Lastly, I thank all the wonderful people and friends I have made along the way from presentations and seminars at conferences (in person and online) from all around the world. Special thanks to Aubrey Poon, Atanas Christev, Gary Koop, Aki Vehtari, Leo Lahti, Bernd Schwaab, Roberto Casarin, Francesco Ravazzolo, Silvia Fruewirth-Schnatter and Herman van-Dijk for giving me the opportunity to present and discuss my research.

Research Thesis Submission

Please note this form should be bound into the submitted thesis.

Name:	David Kohns		
School:	Social Sciences		
Version: <i>(i.e. First, Resubmission, Final)</i>	FINAL	Degree Sought:	PhD

Declaration

In accordance with the appropriate regulations I hereby submit my thesis and I declare that:

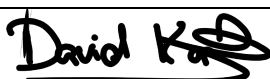
1. The thesis embodies the results of my own work and has been composed by myself
2. Where appropriate, I have made acknowledgement of the work of others
3. The thesis is the correct version for submission and is the same version as any electronic versions submitted*.
4. My thesis for the award referred to, deposited in the Heriot-Watt University Library, should be made available for loan or photocopying and be available via the Institutional Repository, subject to such conditions as the Librarian may require
5. I understand that as a student of the University I am required to abide by the Regulations of the University and to conform to its discipline.
6. I confirm that the thesis has been verified against plagiarism via an approved plagiarism detection application e.g. Turnitin.

ONLY for submissions including published works

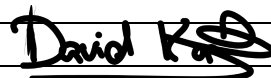
Please note you are only required to complete the Inclusion of Published Works Form (page 2) if your thesis contains published works)

7. Where the thesis contains published outputs under Regulation 6 (9.1.2) or Regulation 43 (9) these are accompanied by a critical review which accurately describes my contribution to the research and, for multi-author outputs, a signed declaration indicating the contribution of each author (complete)
8. Inclusion of published outputs under Regulation 6 (9.1.2) or Regulation 43 (9) shall not constitute plagiarism.

* Please note that it is the responsibility of the candidate to ensure that the correct version of the thesis is submitted.

Signature of Candidate:		Date:	31.01.2023
-------------------------	---	-------	------------

Submission

Submitted By <i>(name in capitals)</i> :	DAVID KOHNS
Signature of Individual Submitting:	
Date Submitted:	31.01.2013

For Completion in the Student Service Centre (SSC)

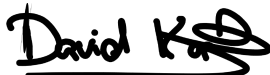
Limited Access	Requested	Yes	No	Approved	Yes	No
<i>E-thesis Submitted (mandatory for final theses)</i>						
Received in the SSC by <i>(name in capitals)</i> :				Date:		

Inclusion of Published Works

Please note you are only required to complete the Inclusion of Published Works Form if your thesis contains published works under Regulation 6 (9.1.2)

Declaration

This thesis contains one or more multi-author published works. In accordance with Regulation 6 (9.1.2) I hereby declare that the contributions of each author to these publications is as follows:

Citation details	Kohns, D., & Bhattacharjee, A. (2022). Nowcasting Growth using Google Trends Data: A Bayesian Structural Time Series Model. International Journal of Forecasting.
Author 1	The idea and topic emerged in much simpler form from my MSc dissertation, and was proposed in its current form by myself. I collected the data, developed the methodology and code, designed and ran the simulations, conducted the empirical analysis and wrote the paper as well as responded to reviewers during the publication process.
Author 2	Prof. Bhattacharjee was instrumental in guiding the writing process as well as editing both the final version of the paper as well the responses to reviewers.
Signature:	
Date:	31.01.2023

Publications

This page lists publications and submitted working papers that were written during my PhD study.

1. Kohns, D. & Bhattacharjee, A. (2022). “Nowcasting Growth using Google Trends Data: A Bayesian Structural Time Series Model”. *International Journal of Forecasting* (forthcoming).
2. Kohns, D. & Potjagailo, G. (2022). “A New Bayesian MIDAS Approach for Flexible and Interpretable Nowcasting”. *Bank of England Working Paper* (forthcoming).
3. Kohns, D. & Szendrei, T. (2021). “Shrink then Sparsify: Identifying Drivers of Tail Risk”. Submitted to *Journal of Econometrics*.
4. Ahrens, A., Aitken, C., Ditzen, J., Ersoy, E., Kohns, D., Schaffer, M. E. (2021). “A Theory-Based Lasso for Time-Series Data”. *Data Science for Financial Econometrics*, pp. 3-36. Springer, Cham., 2021.
5. Kohns, D. & Szendrei, T. (2020). “Horseshoe Prior Bayesian Quantile Regression”. Submitted to *Journal of the Royal Statistical Society: Series A*.

Abstract

This thesis presents new models for nowcasting and macro risk estimation using frontier Bayesian methods that enable incorporating Big Data into policy relevant prediction problems. We propose variable selection algorithms motivated from Bayesian decision theory to make model outcomes interpretable to the policy maker.

In chapter 2, we propose a Bayesian Structural Time Series (BSTS) model for nowcasting GDP growth. This model jointly estimates latent time trends to capture slow moving changes in economic conditions along-side a high dimensional mixed frequency component that is extracted from higher frequency (monthly) cyclical information. We extend on previous implementations of the BSTS with priors and variable selection methods which facilitate selection over latent time trends as well as mixed-frequency information that remain tractable to the policy maker. Empirically, we provide a novel nowcast application where we use a large dimensional set of Internet search terms to gain advance information about supply and demand sentiment for the US economy before more commonly considered macro information are available to the nowcaster. We find that our proposed BSTS model offers large improvements over competing models and that Internet search terms matter for nowcasts before hard information about the macro economy have been published. A simulation exercise confirms the good performance of the proposed model.

Chapter 3 presents the T-SV-t-BMIDAS (Bayesian Mixed Data Sampling) model for nowcasting quarterly GDP growth. The model incorporates a long-run time-varying trend (T) and t-distributed stochastic volatility accounting for outliers (SV-t) into a Bayesian multivariate MIDAS. To address the high-dimensionality of the model, to account for group-correlation in mixed frequency data, and to make the model interpretable to the policy maker, we propose a new combination of group-shrinkage prior with sparsification algorithm for variable selection. The prior flexibly accommodates between-group sparsity and within-group correlation and allows to communicate the joint importance of predictors over the data release cycle. We evaluate the model for UK GDP growth nowcasts covering also the time-span of the Covid-19 recession. The model is competitive prior to the pandemic relative to various benchmark models, while yielding substantial nowcast improvements during the pandemic. Contrary to many previous nowcasting approaches, the model reads in sparse group signals from the data. Simulations show competitive performance of the variable selection methodology, with particularly good performance to be expected for highly correlated data as well as dense data-generating-processes.

Chapter 4 presents a new Bayesian Quantile Regression (BQR) model for high di-

mensional risk estimation. It extends the horseshoe prior to the BQR framework and provides a fast sampling algorithm for computation that makes it efficient for high-dimensional problems. A large scale simulation exercise reveals that compared to alternative shrinkage priors, the proposed methods yield better performance in coefficient bias and forecast error, especially in sparse data-generating processes and in estimating extreme quantiles. In a high dimensional Growth-at-Risk forecasting application, we forecast tail risks as well as complete forecast densities using a database covering over 200 variables related to the U.S. economy. Quantile specific and density calibration score functions show that the horseshoe prior provides the best performance compared to competing Bayesian quantile regression priors, especially at short and medium run horizons.

Bayesian quantile regression models with continuous shrinkage priors are known to predict well but are hard to interpret due to lack of exact posterior sparsity. Chapter 5 bridges this gap by extending the idea of decoupling shrinkage and sparsity. The proposed procedure follows two steps: First, the quantile regression posterior is shrunk via state of the art continuous shrinkage priors; then, the posterior is sparsified by taking the Bayes optimal solution to maximising a policy maker's utility function with joint preference for predictive accuracy as well as sparsity. For the sparsification component, we propose a new variant of the signal adaptive variable selection algorithm that automates the choice of penalization in the integrated utility through a quantile specific loss-function that works well in high dimensions. Large scale simulations show that, compared to the un-sparsified regression posterior, the selection procedure decreases coefficient bias irrespective of the true underlying degree of sparsity in the data, and goodness of variable selection is competitive with traditional variable selection priors. A high dimensional Growth-at-Risk forecasting application to the US shows that the method detects varying degrees of sparsity across the conditional GDP distribution and that the sources to downside risk vary substantially over time.

Inspired by the work of Giannone et al. (2021) on the "illusion of sparsity" from sparse modelling techniques, this chapter (6) investigates whether the recently popularised global-local priors, firstly, are implicitly informative about sparsity and, secondly, whether they are able to communicate the true degree of sparsity from the data. We consider two methods of analysis: implicit model size distributions and sparsification techniques which are tested on a host of economic data sets and simulations. The findings motivate a new horseshoe type model to which we add a prior that makes it a-priori agnostic about the degree of sparsity and is shown to be competitive to the spike-and-slab of Giannone et al. (2021) for forecasting as well as sparsity detection.

Chapter 7 concludes with summaries, limitations of the thesis, as well as directions for future research.

Contents

1	Introduction	1
1.1	Overview	1
1.2	Chapter 2: Nowcasting Growth using Google Trends Data: A Bayesian Structural Time Series Model	5
1.3	Chapter 3: A New Bayesian MIDAS Approach for Flexible and Interpretable Nowcasting	5
1.4	Chapter 4: Horseshoe Prior Bayesian Quantile Regression	6
1.5	Chapter 5: Shrink then Sparsify: Identifying Drivers of Tail Risk	7
1.6	Chapter 6: Global-Local Priors and the Illusion of Sparsity	7
1.7	Chapter 7: Conclusion	8
2	Nowcasting Growth using Google Trends Data: A Bayesian Structural Time Series Model	9
2.1	Introduction	9
2.2	Bayesian Structural Time Series Model	10
2.2.1	The Original Model	10
2.2.2	Model enhancements	11
2.3	Data	16
2.3.1	Mixing Frequencies	16
2.3.2	Macroeconomic Data	17
2.3.3	Google Trends	18
2.3.4	Understanding Google Trends	21
2.3.5	Relationship Between Macroeconomic and Google Search Data	23
2.4	Nowcasting U.S. Real GDP Growth	24
2.4.1	In-Sample Results	26
2.4.2	Nowcast Evaluation: Pre-Pandemic period	30
2.4.3	Nowcast Evaluation: During the Pandemic	33
2.4.4	An Extension to Student-t Errors	34
2.5	Simulation Study	36
2.6	Conclusion	38

3	A New Bayesian MIDAS Approach for Flexible and Interpretable Nowcasting	40
3.1	Introduction	40
3.2	Trend-SV-t Bayesian MIDAS with flexible priors	43
3.2.1	The Model	44
3.2.2	Bayesian Setup	46
3.2.3	Sparsification Step for the GIGG Prior	49
3.3	Data Set and Nowcast Setup	51
3.3.1	Data set	51
3.3.2	Nowcast Exercise	52
3.4	Empirical Results	54
3.4.1	Analysing UK GDP Growth via The Trend-SV-t-BMIDAS	54
3.4.2	Nowcast Evaluation and The Role of Model Features	56
3.4.3	Interpreting Nowcasts via Variable Inclusion Probabilities	62
3.4.4	Sensitivity Analyses	65
3.5	Simulation Study	66
3.5.1	Evaluation	68
3.5.2	Simulation Results	69
3.6	Conclusion	74
4	Horseshoe Prior Bayesian Quantile Regression	76
4.1	Introduction	76
4.2	Methodology	78
4.2.1	Bayesian Quantile Regression	78
4.2.2	Shrinkage Priors	80
4.2.3	Gibbs Sampler	82
4.3	Simulation Setup	83
4.3.1	I.i.d. Distributed Random Error Simulation Results	87
4.3.2	Heteroskedastic Error Simulation Results	91
4.4	Growth at Risk Application	93
4.5	Conclusion	100
5	Shrink then Sparsify: Identifying Drivers of Tail Risk	102
5.1	Introduction	102
5.2	Methodology	105
5.2.1	Bayesian Quantile Regression	105
5.3	Shrinkage Priors	107
5.4	Decoupling Shrinkage and Sparsity	110
5.5	Signal Adaptive Variable Selection for the BQR	112
5.6	Monte Carlo Experiment	115
5.6.1	Monte Carlo Setup	115

5.6.2	Monte Carlo Results	117
5.7	Growth at Risk application	123
5.7.1	Forecast Density Evaluation Results	126
5.7.2	Quantile Specific Variable Inclusion	128
5.8	Conclusion	138
6	Global Local Priors and The Illusion of Sparsity	140
6.1	Introduction	140
6.2	Theory of Model Sizes	143
6.2.1	A General Framework for Implicit Model Sizes	143
6.2.2	Deriving Implicit Model Size Distributions	145
6.2.3	Shrinkage Coefficient Distributions	148
6.2.4	Prior Predictive Model Size Distributions	151
6.2.5	Sparsification Methods	151
6.3	Simulation	157
6.4	Application to Economic Data Sets	162
6.5	Conclusion	165
7	Conclusion	168
A	Appendix of chapter 1	172
A.1	Model Details	172
A.1.1	Posteriors	172
A.1.2	Robustness Check for the Horseshoe Prior	174
A.1.3	ARMA vs LLT Estimation	177
A.1.4	State Space Estimation and Forecasting	179
A.1.5	BSTS-t Model	181
A.2	Graphs	183
A.2.1	HS-BSTS-t Model Results	183
A.2.2	In-Sample Results	184
A.2.3	Google Topic/Category Plots	186
B	Appendix of Chapter 2	187
B.1	Model Details	187
B.1.1	Posteriors	187
B.1.2	Sampling Algorithm	193
B.1.3	Group-Selection Algorithm	193
B.1.4	Almon-Lags Restricted MIDAS	195
B.2	Supplementary Material on the Empirical Application	196
B.2.1	Further Detail on Data	196
B.2.2	In-Sample Description	196
B.2.3	Nowcast Performance GIGG Models	197

B.2.4	Inclusion Probabilities	198
B.2.5	Further Simulation Results	199
C	Appendix of Chapter 3	200
C.1	Derivations and Implementation Details	200
C.1.1	Derivation Algorithm 1	200
C.1.2	Slice Sampling	201
C.2	Supplementary Material on Empirical Application	202
D	Appendix of Chapter 4	214
D.1	Derivations and Further Implimentation Details	214
D.1.1	Expected Loss function for the BQR	214
D.1.2	Derivation of SAVS-BQR	215
D.1.3	Sampling Algorithms	216
D.2	Supplementary Material on the Empirical Application	220
E	Appendix of Chapter 5	222
E.1	Supplementary Material on the Empirical Applications	222
E.1.1	Implicit Model Sizes	222
E.1.2	SAVS Model Sizes	225
E.1.3	Posterior Inclusion Probabilities SAVS	227
E.1.4	ELPD Graphs	231
F	Published Version of Chapter 2	232
F.1	Introduction	232
F.2	Bayesian Structural Time Series Model	233
F.2.1	The Original Model	233
F.2.2	Model enhancements	234
F.3	Data	239
F.3.1	Mixing Frequencies	239
F.3.2	Macroeconomic Data	240
F.3.3	Google Trends	241
F.3.4	Understanding Google Trends	244
F.3.5	Relationship Between Macroeconomic and Google Search Data	246
F.4	Nowcasting U.S. Real GDP Growth	247
F.4.1	In-Sample Results	249
F.4.2	Nowcast Evaluation: Pre-Pandemic period	253
F.4.3	Nowcast Evaluation: During the Pandemic	256
F.4.4	An Extension to Student-t Errors	257
F.5	Simulation Study	259
F.6	Conclusion	261

List of Figures

2.1	PCA of Google Trends	21
2.2	Loadings of GT PCA	23
2.3	Heatmap of correlation	24
2.4	Cumulative one-step-ahead forecast errors in-sample	27
2.5	Posterior inclusion probabilities of the SSVS-BSTS model.	27
2.6	Posterior inclusion probabilities of the HS-SAVS-BSTS model.	27
2.7	Posterior Inclusion probability heatmaps for the SSVS-BSTS	29
2.8	Posterior Inclusion probability heatmaps for the HS-SAVS-BSTS	30
2.9	Distribution of the state standard deviations	30
2.10	Real-Time RMSFE	32
2.11	Real-Time log-predictive density scores	32
2.12	Real-Time CRPS scores	33
2.13	Predictive distributions during the pandemic	34
2.14	Predictive distributions for the full HS-BSTS-t model.	35
3.1	Posterior estimates for Trend-SV-t model components	55
3.2	Posterior trend estimates from different BMIDAS specifications	56
3.3	Nowcast performance for different BMIDAS model components	57
3.4	Nowcast performance of Trend-SV-t-BMIDAS compared to benchmarks	60
3.5	Posterior Nowcasts	62
3.6	Average posterior inclusion probabilities	64
3.7	Robustness checks	65
3.8	Simulation MIDAS lag profiles	68
3.9	Average relative performance over simulations	70
4.1	Distribution of κ_j , the shrinkage coefficient	80
4.2	$\beta_{0,p}$ profiles for DGP y_1	92
4.3	$\beta_{0,p}$ profiles for DGP y_2	92
4.4	β_0 profiles for DGPs y_1 and y_2 in different sparsity settings	93
4.5	β_0 and β_1 profiles for y_3 across quantiles for the different sparsity settings	94
4.6	β_0 and β_1 profiles for y_4 across quantiles for the different sparsity settings	94

4.7	One-step-ahead forecast distributions	97
4.8	Smoothed forecast densities for Q2 2008	100
5.1	Coordinate Descent paths for objective function	114
5.2	Quantile specific sparsity profiles	116
5.3	Cumulative log-scores.	127
5.4	Rossi and Sekhposyan (2019) calibration test.	128
5.5	Inclusion probability heatmaps for HS prior	131
5.6	Inclusion probability heatmaps for HS and LBQR prior	132
5.7	Inclusion probability heatmaps for LBQR and SSVS prior	133
5.8	Correlation coefficients computed on a rolling 30 quarter window . . .	134
5.9	Inclusion probability heatmaps for $HSBQR_{BIC}$ over time	136
5.10	Inclusion probability heatmaps for $LBQR_{BIC}$ and $SSVSBQR$ over time	137
6.1	Shrinkage Distributions for the HS, DL, R2D2, RHS and NG priors, sparse	149
6.2	Shrinkage Distributions for the HS, DL, R2D2, RHS and NG priors, medium sparsity	150
6.3	Shrinkage Distributions for the HS, DL, R2D2, RHS and NG priors, dense	150
6.4	Prior predictive model size distributions for $K = 10$	152
6.5	Prior predictive model size distributions for $K = 100$	153
6.6	Prior predictive model size distributions for $K = 1000$	153
A.1	$\kappa \bullet$ distributions for the horseshoe and the spike-and-slab	175
A.2	Prior predictive of m_{eff} for the horseshoe	176
A.3	Posterior model size distribution for the HS models	176
A.4	Real-Time-RMSFE with HS-ARMA.	177
A.5	Real-Time-LPDS with HS-ARMA.	178
A.6	Real-Time-CRPS with HS-ARMA.	178
A.7	Trend and regression decomposition	179
A.8	Posterior inclusion probabilities for the HS-SAVS-BSTS-t model. . . .	183
A.9	Posterior of the degree of freedom parameter, ν	183
A.10	Predictive distributions for full HS-BSTS-t model	184
A.11	Posterior inclusion probabilities for the original BSTS model of Scott and Varian (2014).	184
A.12	Posterior of the Google category 'Business news 2' and topic 'Invest- ing 0'	185
A.13	U-MIDAS transformed Google search data plots.	186
B.1	Bi-variate shrinkage coefficient plots for GIGG	188

B.2	Posterior Degrees of Freedom of the t-distribution	196
B.3	RSMFE and CRPS for the T-SV-t model with different hyper-parameters	197
B.4	Combination weights of the Combination model	198
C.1	LBQR's β_0 and β_1 profiles for y_3 across quantiles for the different sparsity settings	202
C.2	LBQR's β_0 and β_1 profiles for y_4 across quantiles for the different sparsity settings	202
C.3	Two-step-ahead forecast distributions	203
C.4	Three-step-ahead forecast distributions	203
C.5	Four-step-ahead forecast distributions	203
C.6	One-step-ahead QQ plots of PITs	204
C.7	Two-step-ahead QQ plots of PITs	204
C.8	Three-step-ahead QQ plots of PITs	205
C.9	Fourth-step-ahead QQ plots of PITs	205
C.10	One-step-ahead QQ plots of PITs with only first 50 observations . . .	206
C.11	Forecast densities 1-step Q3 1990	206
C.12	Forecast densities 1-step Q3 2001	207
C.13	Forecast densities 1-step Q1 2005	207
C.14	Forecast densities 2-step Q3 1990	208
C.15	Forecast densities 2-step Q3 2001	208
C.16	Forecast densities 2-step Q3 2008	209
C.17	Forecast densities 2-step Q1 2005	209
C.18	Forecast densities 3-step Q2 1990	210
C.19	Forecast densities 3-step Q4 2001	210
C.20	Forecast densities 3-step Q2 2008	211
C.21	Forecast densities 3-step Q1 2005	211
C.22	Forecast densities 4-step Q3 1990	212
C.23	Forecast densities 4-step Q3 2001	212
C.24	Forecast densities 4-step Q2 2008	213
C.25	Forecast densities 4-step Q1 2005	213
E.1	Macro 1: Implicit model sizes	222
E.2	Macro 2: Implicit model sizes	223
E.3	Finance 1: Implicit model sizes	223
E.4	Micro 1: Implicit model sizes	224
E.5	Macro 2: Implicit model sizes	224
E.6	Macro 1: Sparsified posterior model sizes	225
E.7	Macro 2: Sparsified posterior model sizes	225
E.8	Finance 1: Sparsified posterior model sizes	226
E.9	Micro 1: Sparsified posterior model sizes	226

E.10	Micro 2: Sparsified posterior model sizes	227
E.11	Macro 1: Posterior inclusion probabilities heatmaps, SAVS	227
E.12	Macro 2: Posterior inclusion probabilities heatmaps, SAVS	228
E.13	Finance 1: Posterior inclusion probabilities heatmaps, SAVS	228
E.14	Micro 1: Posterior inclusion probabilities heatmaps, SAVS	228
E.15	Micro 2: Posterior inclusion probabilities heatmaps, SAVS	229
E.16	Macro 1: Posterior variable selection heatmaps, projpred	229
E.17	Macro 2: Posterior variable selection heatmaps, projpred	229
E.18	Finance 1: Posterior variable selection heatmaps, projpred	230
E.19	Micro 1: Posterior variable selection heatmaps, projpred	230
E.20	Micro 1: Posterior variable selection heatmaps, projpred	230
E.21	ELPD scores for model sizes, RHS, Micro 2.	231
E.22	ELPD scores for model sizes, DL, Micro 2.	231
F.1	PCA of Google Trends	244
F.2	Loadings of GT PCA	244
F.3	Heatmap of correlation	247
F.4	Cumulative one-step-ahead forecast errors in-sample	250
F.5	Posterior inclusion probabilities of the SSVS-BSTS model.	250
F.6	Posterior inclusion probabilities of the HS-SAVS-BSTS model.	250
F.7	Posterior Inclusion probability heatmaps for the SSVS-BSTS	252
F.8	Posterior Inclusion probability heatmaps for the HS-SAVS-BSTS	253
F.9	Distribution of the state standard deviations	253
F.10	Real-Time RMSFE	255
F.11	Real-Time log-predictive density scores	255
F.12	Real-Time CRPS scores	256
F.13	Predictive distributions during the pandemic	257
F.14	Predictive distributions for the full HS-BSTS-t model.	258

List of Tables

2.1	Real time calendar	22
2.2	Results of the simulations	37
3.1	Stylised pseudo real-time data release calendar.	52
3.2	Relative Evaluation Metrics	59
3.3	Monte Carlo Simulations: Sparse Coefficient Vector	71
3.4	Monte Carlo Simulations: Dense Coefficient Vector	75
4.1	Summary of simulation setups	86
4.2	Root mean coefficient bias	89
4.3	Root mean squared forecast error	90
4.4	Quantile specific forecast evaluation	98
4.5	Average log scores and MSFE	98
5.1	Summary of simulation setups	118
5.2	Root Mean Coefficient Bias Results	119
5.3	MCC Results	120
5.4	Hit Rate Results	121
5.5	Forecast Evaluation Results	127
5.6	Average Model sizes of the different models	135
6.1	Shrinkage coefficient distributions implied by the GL priors	152
6.2	Simulation Evaluation Metrics	159
6.3	Expected posterior model sizes for applications	163
6.4	RMSFE for the applications	163
B.1	Further Details on Macro Data	196
B.2	Monte Carlo Simulations: DGP 2	199
D.1	Forecast Evaluation Results (Skewed-t distribution)	220
D.2	Nuisance Parameter correction results for the HSBQR	221
F.1	Real time calendar	245
F.2	Results of the simulations	260

Chapter 1

Introduction

1.1 Overview

In recent decades, advances in technology and data collection have made it easier than ever for economists to access large swathes of micro and macro-economic information for which statistical agencies around the world supply easy to use tools for interaction.¹ This is also increasingly true for new unstructured data sources such as internet search terms, satellite data, scanner data, etc. whose appeal stem from being available in near real time as well as offering much greater granularity (Bok et al., 2018) than conventional economic data. The sheer size of the information set also entail challenges for the econometrician, as commonly employed frequentist linear regression models are inappropriate in high dimensions,² both for estimation as well as communication to the policy maker about which data are most important. Particularly in macroeconomic nowcasting and risk estimation, the time-series dimension is usually very short compared to the size of potential covariates and interpretability is made challenging, among other things, by latent time trends, complicated correlation structures, severe economic shocks (such as the ongoing Covid-19 pandemic) and heterogeneous variable impacts across the conditional distribution of the target variable. These fields have policy relevance since key economic data such as GDP have considerable publication lags, for which ‘early’ forecasts (nowcasts) provide up-to-date estimates,³ and risk estimation is a pivotal part of macroprudential policy of many central banks (Board et al., 2011), particularly in wake of the great financial crisis, 2008, which exposed systemic vulnerabilities between financial markets and the real economy (Adrian et al., 2019; Giglio et al., 2016).

¹ See e.g. McCracken and Ng (2020, 2016); Croushore and Stark (2001) for data sources that have majorly influenced macroeconomic data processing and availability.

² By high dimensions we are mostly concerned in this thesis with “fat”, i.e. large covariate dimension, instead of “tall”, i.e. large number of observations, data following the denominations of Varian (2014). In fat data situations, ordinary least squares may, for one, not have a unique solution without assumptions about sparsity in the regression vector Tibshirani (1996), or may produce poor forecasts, due to large variance.

³ Also, nowcast estimates also inform structural economic models (Bańbura et al., 2013).

This thesis contributes to Bayesian methods that allow flexibly addressing these issues in high dimensional nowcast and risk estimation settings while remaining computationally tractable, and, importantly, interpretable to the policymaker.

Although the model structures for the chapters on nowcasting and risk estimation are different, a common contribution is the extension of the recently proposed global-local (GL) priors (see e.g. Polson and Scott (2010); Bhadra et al. (2019) for excellent reviews) for regularisation of large parameter spaces and derivation of interpretable variable importance by taking a Bayesian decision theoretic approach. This overcomes, for one, the problem that the posterior with global-local priors remain non-zero with probability one (Hahn and Carvalho, 2015), and hence can be difficult to infer importance, but also allows to conduct variable selection from a predictive perspective, unlike the more traditionally employed spike-and-slab priors which derive variable importance from marginal likelihoods (Vehtari and Ojanen, 2012). These though can be sensitive to perturbations of the prior (Barbieri and Berger, 2004; O’Hara and Sillanpää, 2009; Fava and Lopes, 2021) and may also not answer the question the policymaker poses when wanting to communicate which model components drive the posterior predictions of the models (Piironen et al., 2020).

The use of shrinkage methods in nowcasting is by no means new. The seminal paper by Schorfheide and Song (2015) introduces the mixed-frequency VAR (MF-VAR) in which contemporaneous and lagged variables are shrunk via the Minnesota prior. The Minnesota prior assumes that the data follow a random walk, with the degree of shrinkage increasing across lags in a deterministic way (Litterman, 1980; Doan et al., 1984). Recent MF-VARs use more flexible global-local priors such as Koop et al. (2020); Gefang et al. (2020). Dynamic factor models (DFM), which find wide use for nowcasting (Giannone et al., 2008; Bańbura et al., 2013), summarise the co-movement of large amounts of macro data via low dimensional factors and have recently (Antolin-Diaz et al., 2017) been extended to model latent time-trends and outlier components (Antolin-Diaz et al., 2021), to address slow moving changes in economic activity and the volatility introduced with the Covid-19 pandemic respectively. Multivariate models such as MF-VARs and DFMs can however be bottlenecked by computational complexity in high-dimensions,⁴ and variable importance can be hard to derive from these without further structural assumptions. Additionally, the Covid-19 shock hit economic sectors in a strongly heterogeneous way which further challenges models that rely on co-movement across many indicators.

Mixed Data Sampling (MIDAS) regression on the other hand are parsimonious

⁴ This is on the one hand due to their multivariate nature, and the other, because they are often estimated with state space methods, where intra low-frequency observations are treated as unknown parameters that require computationally expensive filtering methods (Mariano and Murasawa, 2003).

single equation methods that change their specification depending on data availability and empirically provide excellent performance (Carriero et al., 2019). Previous Bayesian implementations of MIDAS (Carriero et al., 2015) consider standard single equation Minnesota type priors and stochastic volatility, but have not been considered for simultaneous modelling of latent-time trends, global-local priors to regularise large parameter spaces and outlier components.

In chapter 2, we propose a Bayesian Structural Time Series (BSTS) model for nowcasting GDP growth. This model jointly estimates latent time trends to capture slow moving changes in economic conditions along-side a high dimensional MIDAS component extracted from higher frequency cyclical information. We extend on previous implementations of the BSTS with priors that facilitate selection over the form of potential time-trends and extend a frontier global-local prior to the MIDAS component. Empirically, we contribute to the nowcasting literature by including a large dimensional set of Google Trends terms to gain advance information about supply and demand sentiment for the US economy before more commonly considered macro information are available to the nowcaster. In chapter 3, we improve upon this modelling framework, by extending it to fat-tailed stochastic volatility that distinguishes between short-lived and longer-lived error clustering to deal flexibly with differently volatile economic conditions. We propose a new combination of group-prior and group-sparsification method that facilitates group selection and addresses correlation clustering issues in that are present in MIDAS regressions. We conduct a nowcasting exercise on UK data that compares our model to a host of frontier nowcasting approaches. Both chapters provide simulations that investigate finite sample properties in response to variety of mixed frequency data generating processes (DGPs).

Risk estimation in macroeconomics has recently experienced a surge in popularity within the macroeconomic literature with the seminal “Growth-at-Risk” (GaR) study of (Adrian et al., 2019). The premise of GaR is that worsening financial conditions impact the lower tail of the GDP growth distribution more negatively than the median or right tail which is captured via quantile regression (Koenker and Bassett, 1978). Quantile regression allows the regression coefficient vector to vary along the conditional GDP distribution. Quantifying this vulnerability is of key interest to policymakers, as it is well-known that recessions caused by financial crises are often more severe than ordinary recessions (Jurado et al., 2015).

However, the recent Covid-19 pandemic has highlighted that sources of risk can be numerous, and potentially unknown to the modeller. In chapter 4 and 5 we

extend Bayesian shrinkage and sparsification methods to improve risk estimation and interpretability in high dimensions.

The statistical literature prior to publication of chapter 4 has mainly considered lasso type priors for regularisation of high-dimensional Bayesian quantile regression problems (Chen et al., 2013; Wang, 2017; Alhamzawi et al., 2012; Alhamzawi and Yu, 2013). It is, however, well known from the conditional mean literature that exponential-Laplace (lasso) priors tend to over-regularise signals and under-regularise noise (Carvalho et al., 2010). In chapter 4, we extend the global-local shrinkage prior framework with the horseshoe as our focal point to the Bayesian quantile regression. We develop an algorithm that scales well in high-dimensions and contribute to the empirical macro literature by estimating a novel GaR exercise for the US which showcases the performance of the horseshoe prior to the previously proposed lasso priors. We conduct a large scale simulation exercise to investigate how well the proposed algorithm behaves in a variety of DGPs. In chapter 5, we extend the Bayesian decision theoretic variable selection paradigm to the Bayesian quantile regression with continuous shrinkage priors. We derive novel analytical solutions and propose a quantile specific parameter within the algorithm that is data dependent and allows automatic adaptation to the quantile specific degree of sparsity. We use this methodology for a novel growth-at-risk application for the US in which we investigate drivers of GDP risk across quantiles and across time.

The analyses based on global-local priors presented in chapters 2-5 lay out new findings about the debate on the “illusion of sparsity” initiated by Giannone et al. (2021). The authors find that the degree of sparsity found in models which are not able to nest dense model solutions may falsely find evidence for sparsity. Since this has an impact on how we interpret the data and may also influence subsequent modelling in the larger Bayesian workflow (Gelman et al., 2020), this is a debate closely followed also in policymaking. A drawback of the original and related studies is that they consider spike-and-slab priors which may induce their own sparsity artefacts due to discretisation of the model space and the strict assumption of a normal slab distribution (Fava and Lopes, 2021). In chapter 6, we take a bird’s eye view on the GL priors used in the previous chapters, as well as other popular global-local priors, and analyse, on one the one hand, whether they have a-priori tendencies about the degree of sparsity, and on the other, whether they are able to communicate the true sparsity once they have seen the data. We consider two methods of analysis: implicit model size distributions which are derived following Piironen et al. (2017), and popular sparsification techniques which are tested on a host of economic data sets and simulations. The findings motivate a new horseshoe type model to which we add a prior that makes it a-priori agnostic about the degree of sparsity.

1.2 Chapter 2: Nowcasting Growth using Google Trends Data: A Bayesian Structural Time Series Model

This chapter 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, originally proposed by Scott and Varian (2014), to make these better amenable to nowcasting with large number of potential covariates and heterogeneous time trends. Specifically, we allow shrinking state variances towards zero to avoid overfitting latent trends, extend the SSVS (spike and slab variable selection) prior to the more flexible normal-inverse-gamma prior (Ishwaran et al., 2005) which stays agnostic about the underlying model size, as well as adapt the horseshoe prior (Carvalho et al., 2010) 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 Internet search terms is able to improve nowcasts early in a given 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. Focusing on the months of the recession induced by Covid-19, inclusion of Google Trends allows additionally for better calibrated forecast densities. Here the BSTS models show a large asymmetric dip in the lower quantiles of the nowcast distribution, anticipating the trough of the recession. This is not captured when excluding the search terms from the information set.

1.3 Chapter 3: A New Bayesian MIDAS Approach for Flexible and Interpretable Nowcasting

In this chapter we build on the work of chapter 2 by proposing the T-SV-t-BMIDAS model for nowcasting quarterly GDP growth. The model incorporates a long-run time-varying trend (T) and t-distributed stochastic volatility accounting for outliers (SV-t) into a Bayesian multivariate MIDAS. To address the high-dimensionality of the model, to account for group-correlation in mixed frequency data, and to make the model interpretable to the policy maker, we propose a new combination of group-shrinkage prior and sparsification algorithm for variable selection in the MIDAS component. The prior offers three-tiered shrinkage in which the overall level of sparsity, group-level sparsity and the degree of correlation between group members

are controlled by individual hyper-parameters. This addresses, on the one-hand, the grouping structure implied by MIDAS transformed mixed frequency lags, and the other, distinguishes between situations where the impact of within group lags are correlated or heterogeneous. This also helps avoiding mixing issues often faced by shrinkage priors for highly correlated groups (Pironen et al., 2020). To communicate variable impact on model predictions, we derive group-wise inclusion probabilities following the Bayesian decision theoretic framework of Hahn and Carvalho (2015) and Chakraborty et al. (2020). We evaluate the model for UK GDP growth nowcasts over the period 1999 to 2021. Our model is competitive prior to the Covid-19 pandemic relative to various benchmark models, while yielding substantial nowcast improvements during the pandemic. First, accounting for a long-run trend and t-distributed stochastic volatility substantially improves forecast performance relative to a simple BMIDAS. Second, the shrinkage prior enhances nowcast performance by inducing group-wise sparsity while enabling the model to flexibly shift between signals. During the Covid-19 pandemic, the model reads stronger signals from indicators for services, which reflected spending shifts related to lockdowns, and less from production surveys. This helps to precisely nowcast the initial recovery after the shock, and to update the nowcast for the pandemic-related trough sooner. Simulations show competitive performance of the shrinkage and variable selection methodology, with particularly good performance to be expected for highly correlated data as well as dense data-generating-processes compared to the horseshoe prior and a frontier spike-and-slab based group-prior.

1.4 Chapter 4: Horseshoe Prior Bayesian Quantile Regression

This chapter marries the burgeoning literature of high-dimensional global-local shrinkage priors with the Bayesian quantile regression and provides an efficient sampling algorithm for high-dimensional quantile regression problems. With the horseshoe prior as the focal point, we provide a competitive prior that compares favourably to a host of previously proposed lasso based shrinkage priors for the Bayesian quantile regression. Empirically, this chapter highlights the performance of the proposed method through a novel big-data Growth at Risk application to the US economy in response to 230 potential variables. The premise of this application is that risks to the macroeconomy may stem from many, potentially a-priori unknown, macro sources which the prior can exploit for superior downside risk and general density estimation. We show in our application that the proposed prior provides policy relevant characterisations of downside risks over time.

Finally, we validate the good results from the empirical application in a large-scale simulation exercise that compares the horseshoe’s performance to a host of

competing shrinkage priors in various dense and sparse data generating processes.

1.5 Chapter 5: Shrink then Sparsify: Identifying Drivers of Tail Risk

Bayesian quantile regression models with continuous shrinkage priors are known to predict well but are hard to interpret due to lack of exact posterior sparsity. This chapter bridges this gap by extending the idea of decoupling shrinkage and sparsity in vain of Lindley (1968) and Hahn and Carvalho (2015). The proposed procedure follows two steps: First, we shrink the quantile regression posterior through state of the art continuous shrinkage priors; then, we sparsify the posterior by taking the Bayes optimal solution to maximizing a policy maker’s utility function that considers the predictive performance of the un-sparsified model as well as sparsity. To conduct variable selection, we propose a new variant of the signal adaptive variable selection algorithm (Ray and Bhattacharya, 2018) that automates the choice of penalization in the integrated utility through quantile specific loss-functions that work well in high dimensions. We show in large scale simulations that, compared to the un-sparsified regression posterior, our selection procedure decreases coefficient bias irrespective of the true underlying degree of sparsity in the data, and goodness of variable selection is competitive with traditional variable selection priors. We apply our approach to a high dimensional growth-at-risk exercise in which we identify and communicate to the policy maker which variables drive tail risks to the macroeconomy.

1.6 Chapter 6: Global-Local Priors and the Illusion of Sparsity

Inspired by the work of Giannone et al. (2021) on the illusion of sparsity from sparse modelling techniques, we investigate in this chapter whether the recently popularised global-local priors are able to uncover the true degree of sparsity in the data via an implicit measure of sparsity derived from the model as well as sparsity implied from post-estimation variable selection algorithms which are gaining popularity for communicating the posterior of continuous shrinkage priors (Ray and Bhattacharya, 2018; Huber et al., 2019; Piironen et al., 2020). The analysis motivates a new variant of the horseshoe prior of Carvalho et al. (2010) that a-priori stays agnostic about model complexity. The simulations as well as empirical applications show that the proposed prior not only compares favourably to the spike-and-slab prior of Giannone et al. (2021) in terms of forecasting performance, but is also competitive in pinning down the true degree of sparsity. The analysis also suggest that certain classes of global-local priors are more robust at detecting the underlying degree of sparsity

which we can relate to a prior's implicit distributions on expected model sizes. We find that sparsification methods cannot make up the prior's a-priori sparsity tendencies, but that incorporating out-of-sample predictive performance can somewhat remedy variable selection particularly in dense data generating processes.

1.7 Chapter 7: Conclusion

This chapter summarises the thesis, highlights the contributions made in terms of methodology as well as empirics. We end the thesis with the limitations of the presented research and avenues for future research.

Chapter 2

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

2.1 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 chapter 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

¹ The 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.

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 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 2.3, 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 2.4, we present results based on our empirical application of nowcasting U.S. GDP growth, which is followed in section 2.5 by the results from our simulation study. Finally, Section 2.6 concludes with a discussion and avenues for future research.

2.2 Bayesian Structural Time Series Model

2.2.1 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:

$$\begin{aligned}
y_t &= \tau_t + x_t' \beta + \delta_t + \epsilon_t, \epsilon_t \sim N(0, \sigma_y^2) \\
\tau_t &= \tau_{t-1} + \alpha_t + \epsilon_t^\tau, \epsilon_t^\tau \sim N(0, \sigma_\tau^2) \\
\alpha_t &= \alpha_{t-1} + \epsilon_t^\alpha, \epsilon_t^\alpha \sim N(0, \sigma_\alpha^2) \\
\delta_t &= - \sum_{s=1}^{S-1} \delta_{t-s} + \epsilon_t^\delta, \epsilon_t^\delta \sim N(0, \sigma_\delta^2).
\end{aligned} \tag{2.1}$$

(2.1) 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 ($N\mathcal{G}^{-1}$) 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.

2.2.2 Model enhancements

Non-Centred Bayesian Structural Time Series

In line with previous nowcasting studies, this chapter focuses on nowcasting GDP *growth* rather than levels. However, two problems arise when applying model (2.1) 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

² Modelling 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 (2012a).

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 \mathcal{G}^{-1} priors with no prior mass on zero, as implemented for Bayesian linear state space methods, can bias posterior state variances away from zero, thereby potential leading to false support for state dynamics which can hurt forecast performance.

We extend model (2.1) to flexibly let the data shut down state dynamics, and therefore broaden the applicability of model (2.1), 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 \mathcal{G}^{-1} 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 2.4. Testing for zero posterior variance would be very challenging in a frequentist hypothesis testing approach because the null hypothesis of constant state in model (2.1) lies on the boundary of the parameter space.

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

$$y_t = \tau_0 + \sigma_\tau \tilde{\tau}_t + t\alpha_0 + \sigma_\alpha \sum_{s=1}^t \tilde{\alpha}_s + x_t' \beta + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma^2) \quad (2.2)$$

and

$$\begin{aligned} \tilde{\tau}_t &= \tilde{\tau}_{t-1} + \tilde{u}_t^\tau, \quad \tilde{u}_t^\tau \sim N(0, 1) \\ \tilde{\alpha}_t &= \tilde{\alpha}_{t-1} + \tilde{u}_t^\alpha, \quad \tilde{u}_t^\alpha \sim N(0, 1) \end{aligned} \quad (2.3)$$

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 (2.2) and (2.3) is equivalent to (2.1), let:

$$\begin{aligned} \alpha_t &= \alpha_0 + \sigma_\alpha \tilde{\alpha}_t \\ \tau_t &= \tau_0 + \sigma_\tau \tilde{\tau}_t + t\alpha_0 + \sigma_\alpha \sum_{s=1}^t \alpha_s \end{aligned} \quad (2.4)$$

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

⁴ For further discussion, please see section 2.3

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

$$\begin{aligned}
\alpha_t - \alpha_{t-1} &= \sigma \alpha (\tilde{\alpha}_t - \tilde{\alpha}_{t-1}) \\
&= \sigma_\alpha + \tilde{u}_t^\alpha \\
\tau_t - \tau_{t-1} &= \alpha_0 + \sigma_\alpha \tilde{\alpha}_t + \sigma_\tau (\tilde{\tau}_t - \tilde{\tau}_{t-1}) \\
&= \alpha + \sigma_\tau + \tilde{u}_t^\tau
\end{aligned} \tag{2.5}$$

which recovers (2.1). 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, 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:

$$\theta \sim N(\theta_0, V_\theta). \tag{2.6}$$

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.1.4) along with the state posteriors. In contrast to FFBS type algorithms, it samples the states without recursive estimation which speeds up computation significantly.

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:

$$\begin{aligned}
\beta_j | \gamma_j, \delta_j^2 &\sim \gamma_j N(0, \delta_j^2) + (1 - \gamma_j) N(0, c \times \delta_j^2) \\
\delta_j^2 &\sim \mathcal{G}(a_1, a_2) \\
\gamma_j &\sim \text{Ber}(\pi_0) \\
\pi_0 &\sim \mathcal{B}(b_1, b_2)
\end{aligned} \tag{2.7}$$

where $j \in (1, \dots, K)$ and \mathcal{G} , Ber and \mathcal{B} denote the Gamma, Bernoulli and Beta distribution respectively. 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 through 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, (2.7) 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.1.1). 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.

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:

$$\begin{aligned}
 \beta_j | \varphi_j^2, \vartheta^2, \sigma^2 &\sim N(0, \varphi_j^2 \vartheta^2 \sigma^2), j \in (1, \dots, K) \\
 \varphi_j &\sim \pi(\varphi_j), j \in (1, \dots, K) \\
 \vartheta &\sim \pi(\vartheta) \\
 \sigma^2 &\sim \pi(\sigma^2)
 \end{aligned} \tag{2.8}$$

The idea of this family of priors is that the global scale ϑ controls the overall shrinkage applied to the regression, while the local scale φ_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 ϑ and φ_j :

$$\begin{aligned}
 \varphi_j &\sim C_+(0, 1) \\
 \vartheta &\sim C_+(0, 1)
 \end{aligned} \tag{2.9}$$

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 (Piiironen et al., 2017) (see section A.1.2 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 (2.9).

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 (Piiironen et al., 2017).⁵ We provide in the appendix a (A.1.2) robustness check based on Piiironen et al. (2017) that are able to alleviate any identifiability concerns for the empirical study below.

Posteriors are described in the appendix (A.1.1).

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:

$$\psi_j = \text{sign}(\hat{\beta}_j) \|X_j\|^{-2} \max(|\hat{\beta}_j| \|X_j\| - \zeta_j, 0), \quad (2.10)$$

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 ζ_j in (2.10) 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

$$\zeta_j = \frac{1}{|\hat{\beta}_j|^2}, \quad (2.11)$$

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 similarity to lasso style regularisation, the solution to (2.10)

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

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

$$\bar{\psi} = \underset{\psi}{\operatorname{argmin}} \left\{ \frac{1}{2} \|X\hat{\beta} - X\psi\|^2 + \sum_{j=1}^K \zeta_j |\psi_j| \right\}. \quad (2.12)$$

Here, $\bar{\psi}$ 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).

Sampling Algorithm

With the conditional posteriors for the regression and state components at hand (see Appendix A.1.1), 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 2.2.2, 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.1.4. As discussed in Section 2.2.2, 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 (2.10) on an iteration basis.

2.3 Data

2.3.1 Mixing Frequencies

In this chapter, 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 (Bańbura 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 (2.1) 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 β . This vertical realignment is visualised for a single representative regressor below:

$$\left(\begin{array}{c|ccc} y_{1stquarter} & x_{Mar} & x_{Feb} & x_{Jan} \\ y_{2ndquarter} & x_{Jun} & x_{May} & x_{Apr} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \end{array} \right) \quad (2.13)$$

2.3.2 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 data set.^{6,7}

As early data vintages of macroeconomic data and GDP figures can exhibit sub-

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

⁷ We 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 chapter remain qualitatively identical. Details are available upon request.

stantial 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 (2.1) 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

2.3.3 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 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

⁸ The 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 evaluation.

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

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

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

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
- **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

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

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 A.13 in appendix). Topics around housing have been shown to be indicative of housing prices (Wu and Brynjolfsson, 2015; Askitas 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., 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

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

¹⁴ To 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.

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 2.1 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 2.4. 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.

2.3.4 Understanding Google Trends

To get a visual understanding of the Google search information, we plot in figure (2.1) the first 3 principal components of the U-MIDAS transformed Google Trends information. Figure (2.2) shows the factor loadings.

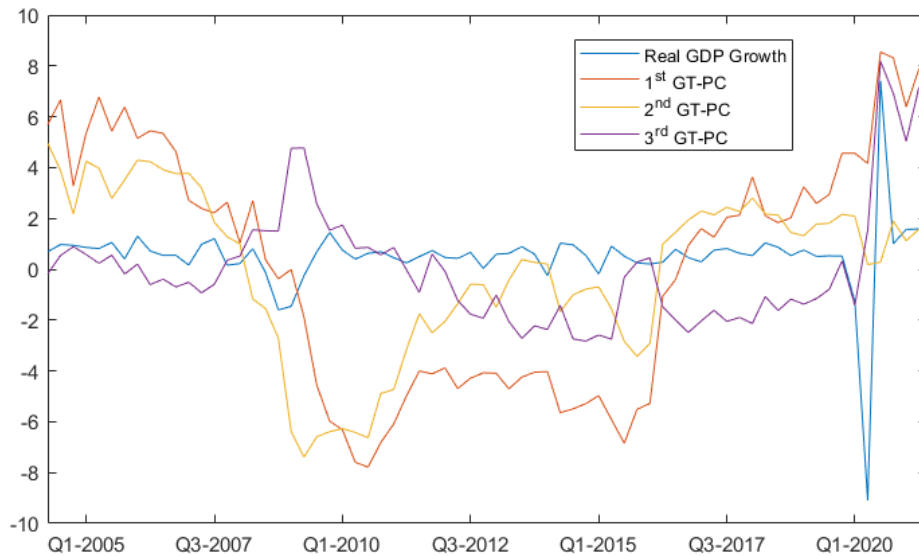


Figure 2.1: First 3 principal components of the GT data set and 2nd vintage real GDP growth.

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 A.13 for indicative time-series plots of individual

¹⁵ Due 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	USEPUINDEXM
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	USEPUINDEXM
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	USEPUINDEXM
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 2.1: 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.

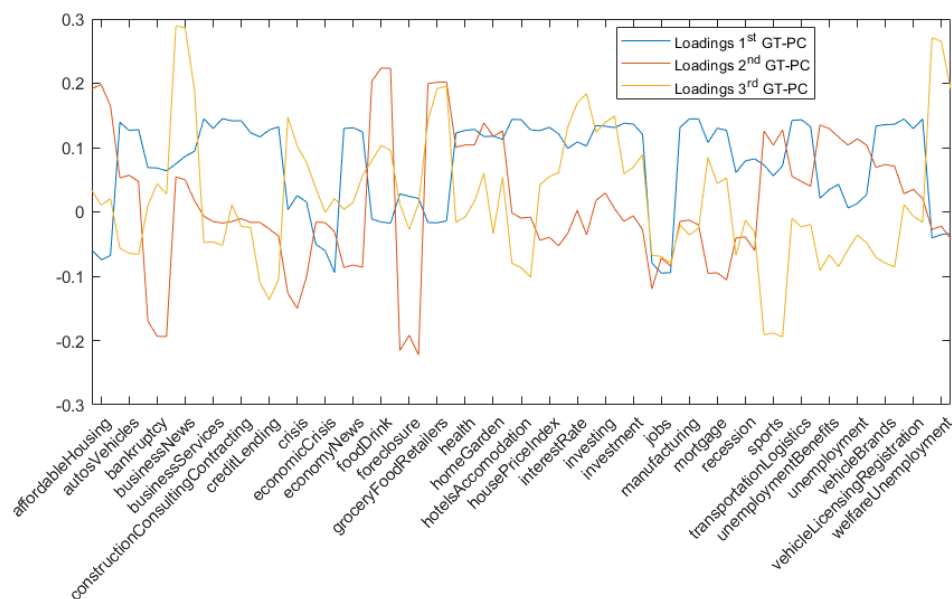


Figure 2.2: Loadings for first 3 principal components of the GT data set.

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 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.

2.3.5 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 (2.3) shows a correlation heatmap between the macro data set and the first three Google Trends principal components. Please note that an increase in the UMPI and PMI indicate improved consumer and producer sentiments,

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

respectively.

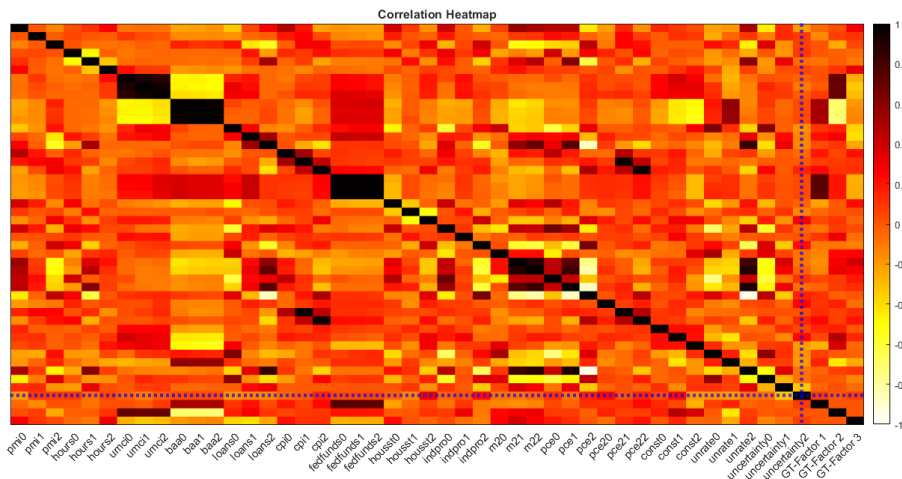


Figure 2.3: Heatmap of sample correlation matrix.

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 consumer 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.

2.4 Nowcasting U.S. Real GDP Growth

The predictive model used to generate in-sample and out-of-sample predictions is given in equations (2.2)-(2.3) where the first $T = 45$ observations are used as training sample.¹⁷ We estimate three variants of the model based on priors (2.7, 2.9, and 2.10) 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

¹⁷ As 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.

growth observation y_{T+1} , conditional on the real-time information set Ω_T^v , where ($v = 1, \dots, 37$) refers to nowcast periods within the real-time calendar (Table 2.1). 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:

$$\text{RT-RMSFE} = \sqrt{\frac{1}{T_{end}} \sum_{j=1}^{T_{end}} (y_{T+j} - \hat{y}_{T+j|\Omega_{T+j-1}^v}^v)^2}, \quad (2.14)$$

where $\hat{y}_{T+j|\Omega_{T+j-1}^v}^v$ 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):

$$\begin{aligned} \text{RT-LPDS} &= \frac{1}{T_{end}} \sum_{j=1}^{T_{end}} \log p(y_{T+j} | \Omega_{T+j-1}^v) \\ &= \frac{1}{T_{end}} \sum_{j=1}^{T_{end}} \log \int p(y_{T+j} | \Omega_{T+j-1}^v, \Upsilon_{1:T+j-1}) p(\Upsilon_{1:T+j-1} | \Omega_{T+j-1}^v) d\Upsilon_{1:T+j-1} \\ &\approx \frac{1}{T_{end}} \sum_{j=1}^{T_{end}} \log \left(\frac{1}{M} \sum_{m=1}^M p(y_{T+j} | \Omega_{T+j-1}^v, \Upsilon_{1:T+j-1}^m) \right), \end{aligned} \quad (2.15)$$

$$\text{RT-LPDS} = \frac{1}{T_{end}} \sum_{j=1}^{T_{end}} \frac{1}{2} \left| y_{T+j} - y_{T+j|\Omega_{T+j-1}^v}^v \right| - \frac{1}{2} \left| y_{T+j|\Omega_{T+j-1}^v}^{v,A} - y_{T+j|\Omega_{T+j-1}^v}^{v,B} \right|, \quad (2.16)$$

where, for brevity of notation, $\Upsilon_{1:T+j-1}$ 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 (2.16), $y_{T+j|\Omega_{T+j-1}^v}^{v,A,B}$ are independently drawn from the posterior predictive density $p(y_{T+1}^v | \Omega_{T+j-1}^v | y_T)$.

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 (2.14, 2.15, and 2.16) 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.1.4.

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

$$\text{DS} = \frac{p(\sigma^{\tau,\alpha} = 0)}{p(\sigma^{\tau,\alpha} = 0|y)} \quad (2.17)$$

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.

2.4.1 In-Sample Results

Figure 2.4 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 2.4, 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 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.

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

¹⁹ Although 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.

²⁰ The 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.

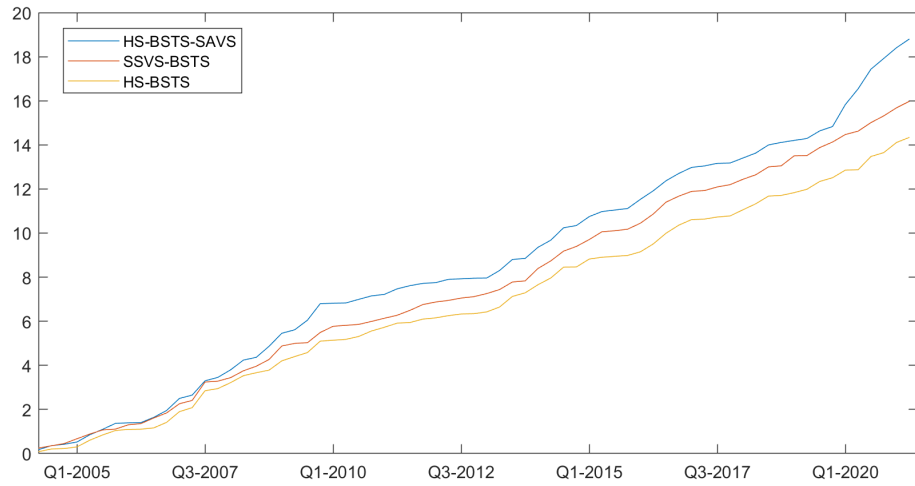


Figure 2.4: Cumulative one-step-ahead forecast errors in-sample from 3 different models: (1) SSVS-BSTS, (2) HS-BSTS and (3) HS-BSTS-SAVS

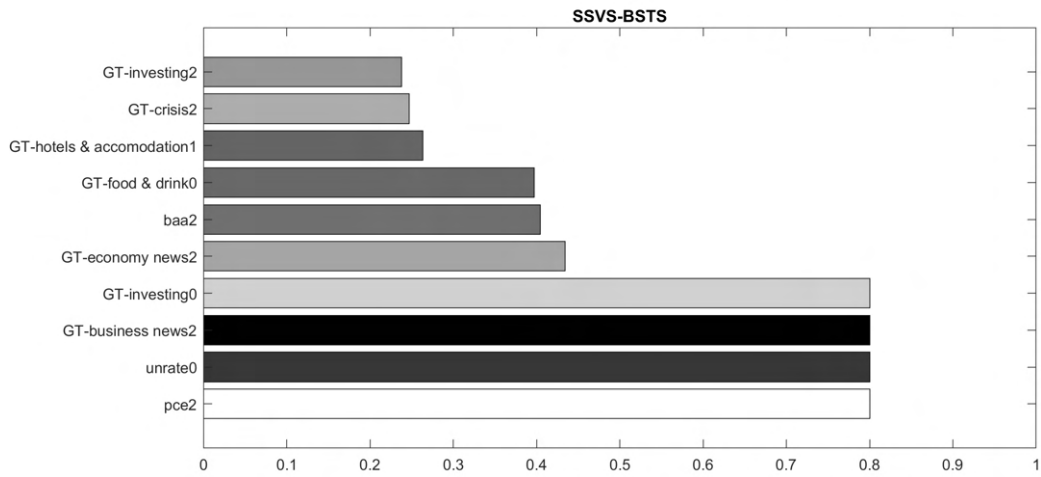


Figure 2.5: Posterior inclusion probabilities of the SSVS-BSTS model.

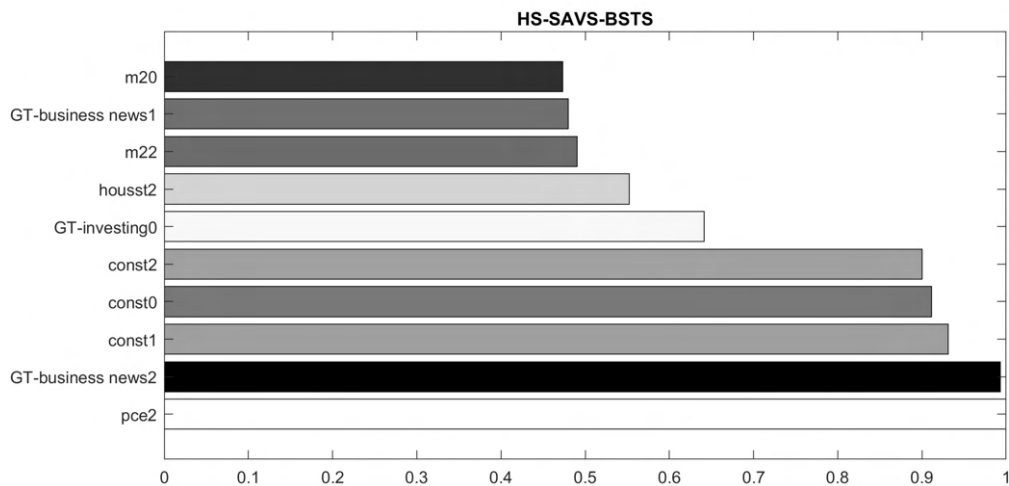


Figure 2.6: Posterior inclusion probabilities of the HS-SAVS-BSTS model.

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 2.5 and 2.6 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 A.11), 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 A.12) 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., 2020; 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 A.13, 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 2.5 and 2.6 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

²¹ It 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.

highest marginal likelihood. Continuous shrinkage priors, on the other hand, make use of all covariates since they are always included in the model.

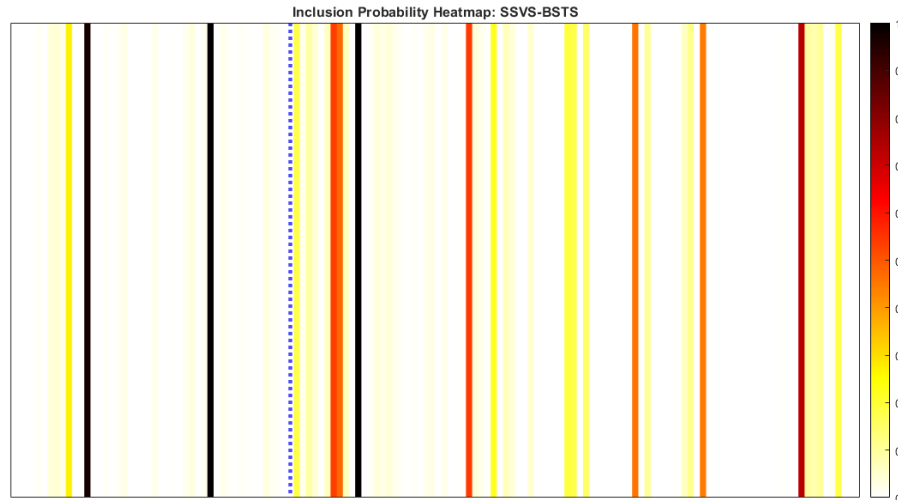


Figure 2.7: Posterior Inclusion probability heatmaps for the SSVS-BSTS. Inclusion probabilities to the left of the dashed line pertain to the macro data set, and to the right, the Google search data.

In line with this discussion, the posterior inclusion probability heatmaps in Figures 2.7 and 2.8 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 and Google Trends) and within mixed frequency information. And secondly, different sparsity patterns can emerge depending on the prior used. Section 2.5 further investigates robustness of the proposed priors to different sparsity settings, and provide recommendations.

Finally, our in-sample results clearly demonstrate (Figure 2.9) that there is support in the data for a local trend, but not a local linear trend: the posterior for σ^τ

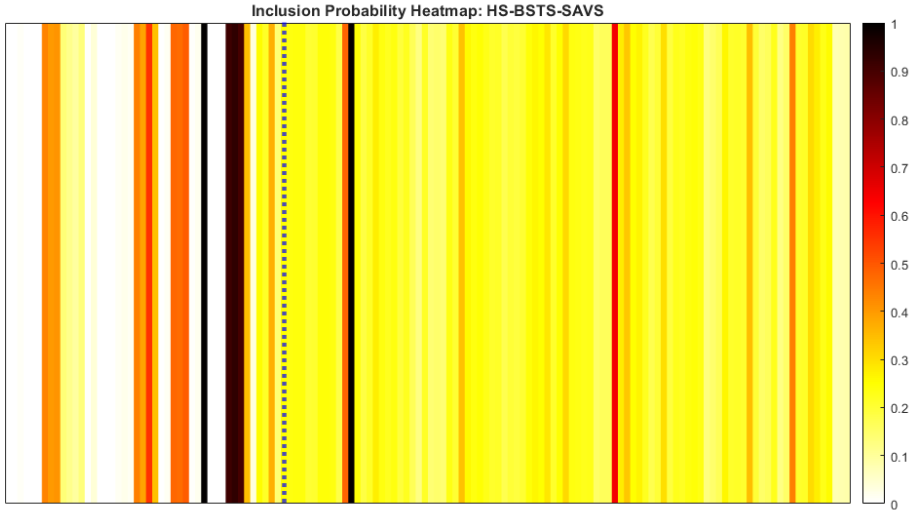


Figure 2.8: Posterior Inclusion probability heatmaps for the HS-SAVS-BSTS. Inclusion probabilities to the left of the dashed line pertain to the macro data set, and to the right, the Google search data.

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.

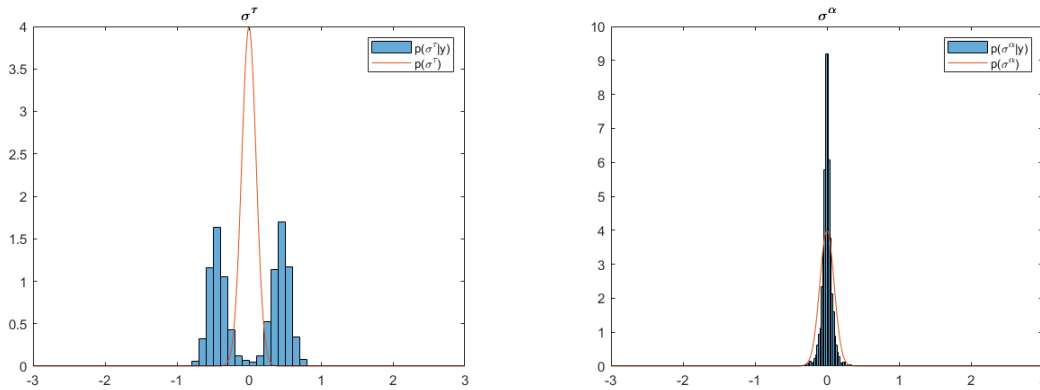


Figure 2.9: Distribution of the (left) trend state standard deviation and (right) slope standard deviation for the HS-BSTS model.

2.4.2 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 2.3. 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 2.10 for the competing non-centred BSTS esti-

mators, 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 2.10. 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 consequently poor nowcasting performance.

²² The 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.

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

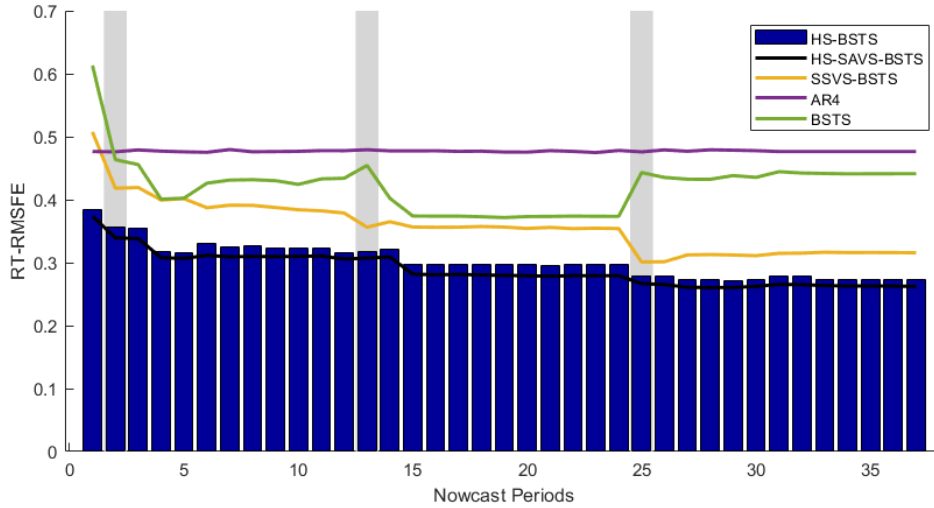


Figure 2.10: Real-Time RMSFE of all competing models. The RT-RMSFE for the BSTS are plotted on the right axis. Grey vertical bars indicate nowcast periods in which GT are published.

Similar to the real-time point forecasts, we plot real-time LPDS (RT-LPDS) and CRPS (RT-CRPS) in Figures 2.11 and 2.12, 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.

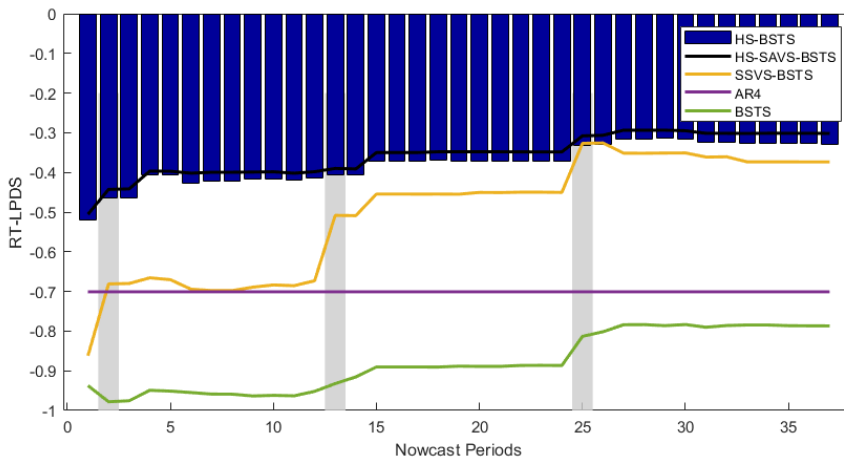


Figure 2.11: Real-Time log-predictive density scores (RT-LPDS) for all competing models. The RT-LPDS for the BSTS are plotted on the right axis. Grey vertical bars indicate nowcast periods in which GT are published.

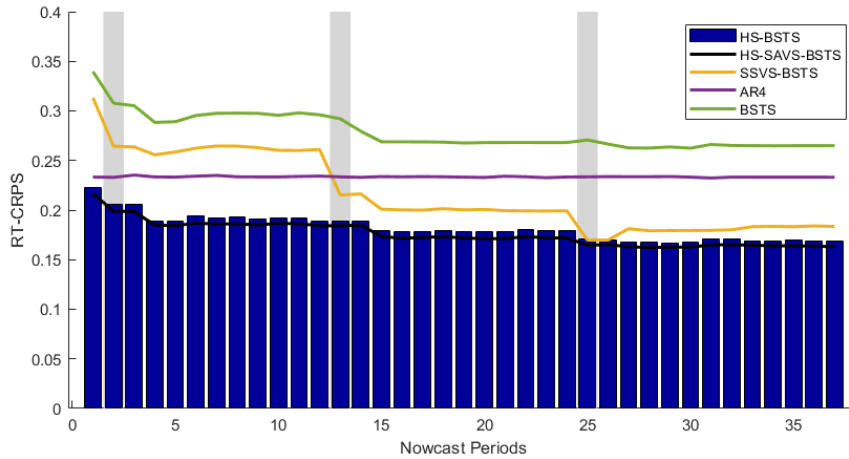


Figure 2.12: Real-Time CRPS scores (RT-CRPS) for all competing models. The RT-CRPS for the BSTS are plotted on the right axis. Grey vertical bars indicate nowcast periods in which GT are published.

2.4.3 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 2.13 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

²⁴ Higher 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.

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

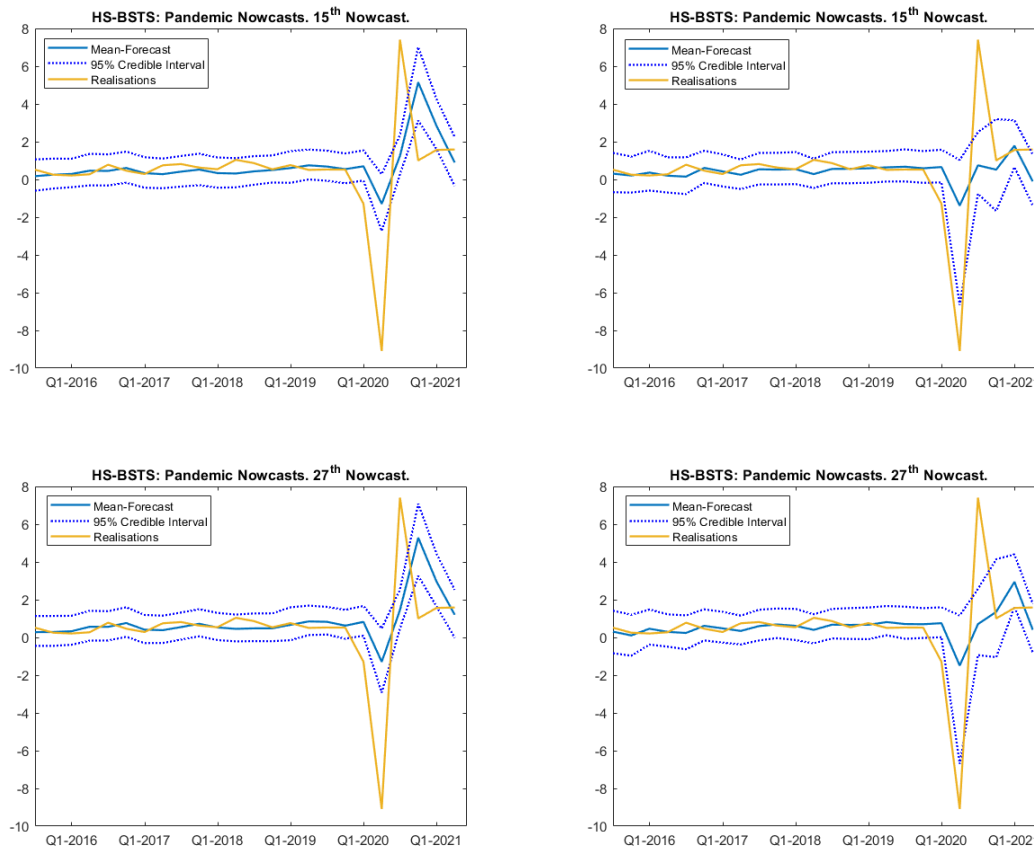


Figure 2.13: Predictive distributions for the macro only data set (left column) and the macro + Google Trends data set (right columns) for the 15th (upper row) and 27th (lower row) nowcast period.

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.

2.4.4 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 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

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

down-weighting outliers through the error covariances.

The estimated model uses the same regression and state components as model (2.2)-(2.3). However, we assume $\epsilon_t \sim N(0, \sigma^2 \lambda_t)$, where λ_t is distributed as $\mathcal{G}^{-1}(\nu/2, \nu/2)$ where \mathcal{G}^{-1} denotes the inverse-Gamma distribution and ν the degree of freedom parameter of the t-distribution. 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.1.5). Treating ν as a random variable, figure (A.9), 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.

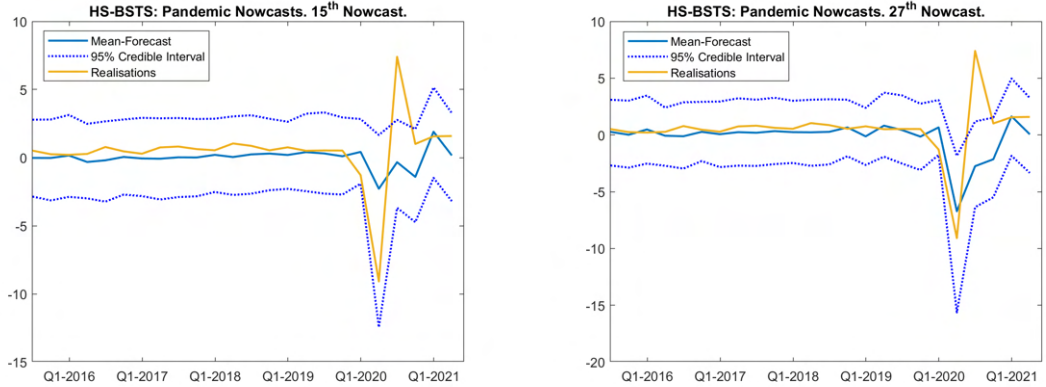


Figure 2.14: Predictive distributions for the full HS-BSTS-t model.

The nowcasts from this model (Figure 2.14) 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 (A.10) 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 (A.8) 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.

2.5 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 chapter 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 (2.2)-(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

$$\beta_{sparse} = (1, 1/2, 1/3, 1/4, 1/5, 0_{K-5})' \quad (2.18)$$

and the dense coefficient vector is

$$\beta_{dense} = \begin{cases} 1/3 & \text{with probability } p_d \\ 0 & \text{with probability } 1 - p_d \end{cases}, \quad (2.19)$$

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 (2.13). 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

$$\text{Root Mean Coefficient Bias} = \sqrt{\frac{1}{20} \|\hat{\beta} - \beta\|_2^2}, \quad (2.20)$$

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.

$(\sigma^\tau, \sigma^\alpha)$	Sparse				Dense			
	(0.5,0)	(0,0.5)	(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 2.2: 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 2.2, 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

²⁷ Note that we do not report DS tests for the original BSTS model. This is due to the fact that the prior on the state variance has no mass on zero and therefore is not testable.

2.6 Conclusion

In this chapter, 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 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.

Chapter 3

A New Bayesian MIDAS Approach for Flexible and Interpretable Nowcasting

3.1 Introduction

The Covid-19 pandemic has raised new challenges for nowcasting GDP. The large size of the shock caused parameter instabilities in time series methods and has challenged the assumption of a constant variance or slow-moving stochastic volatility in the conditional GDP growth distribution. Additionally, the Covid-19 shock hit economic sectors in a strongly heterogeneous way such that standard survey indicators typically employed for nowcasting provided only weak signals for economic activity, which further challenged models that rely on co-movement across many such indicators. More generally, this episode has illustrated the need for more flexible model features that allow for robust and interpretable nowcasts in presence of large shocks, heterogeneous signals, and regime shifts. This remains relevant given the continuing large economic distortions following the pandemic, Russia's invasion of Ukraine, and the high inflation rates observed across advanced economies.

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

On the other hand, Mixed-data sampling (MIDAS) regressions (Ghysels et al.,

2007) have remained under-exploited in nowcasting GDP since the pandemic, despite their competitive nowcast performance in the past (Clements and Galvão, 2009; Foroni et al., 2015; Carriero et al., 2019) and their popularity in policy institutions (Anesti et al., 2017). Unlike factor models, MIDAS models do not rely on co-movement in data but instead include higher-frequency indicators and their lags individually. This can be desirable to exploit heterogeneous signals from the data flexibly during a large shock that affects economic sectors differently. However, multivariate MIDAS models are also prone to various issues: First, such models are highly parameterised, and therefore prone to overfitting without efficient shrinkage. This also makes it challenging to extend such models to non-linear and flexible model features. Second, the typically large group-correlation between high-frequency lags can cause mixing problems for aggressive shrinkage priors (Piiironen et al., 2020; Ishwaran et al., 2005; Malsiner-Walli and Wagner, 2018) which can result in models randomly loading onto certain lags within a group. This may hurt both nowcast performance and interpretability of nowcast signals.

In this chapter, we propose the Trend-SV-t BMIDAS for flexibly nowcasting GDP before and during the Covid-19 pandemic. We incorporate extreme observations and shifts over time via a time-varying trend and flexible outlier components into a multivariate MIDAS framework. The model comprises 1) a cyclical component that captures the responses of GDP to a set of higher frequency indicators, 2) a time varying trend component to capture low-frequency changes in GDP growth and 3) stochastic volatility processes that account for fat tails in the conditional GDP growth distribution. These model features make the model more flexible in dealing with gradual shifts in long-run GDP growth that, if unaccounted for, could blur signals from high-frequency indicators, as well as with large and sudden shifts in the volatility in the data, as observed during the pandemic.

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

In order to communicate variable importance for the nowcasts from the MIDAS posterior, we propose to combine the GIGG prior with a new sparsification algo-

rithm for group-wise variable selection motivated by Bayesian decision theory. The algorithm finds those high-frequency lag groups that best summarise the predictions of the model in the spirit of Hahn and Carvalho (2015) and Chakraborty et al. (2020) and shrinks unimportant high-frequency indicators to exact zeros, thus allowing to derive inclusion probabilities at each point in time. This allows us to communicate signals drawn upon over time and across groups of predictors, thus enhancing the interpretability of nowcasts.

In an empirical application of the Trend-SVt-BMIDAS model, we nowcast UK quarter-on-quarter GDP growth using a set of monthly macroeconomic indicators. We evaluate nowcasts over the data release cycle in a pseudo-real-time setting over the sample period 1999 to 2021, distinguishing between a pre-pandemic sample and the full sample including the Covid-19 shock.

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

Second, the proposed model provides nowcast gains against simpler specifications without time-varying components and against frontier nowcasting models, where our model performs noticeably better during the Covid-19 pandemic. The results show that accounting either for a time-varying trend or for SV with t-errors substantially improves the nowcasts, and that accounting for both provides further gains, particularly pre-pandemic. Intuitively, the inclusion of t-errors helps discount noisy observations for each of the model components. We also find that the Trend-SV-t BMIDAS model is competitive against frontier nowcasting benchmark models such as Combined MIDAS (Mazzi et al., 2014) and a DFM (Antolin-Diaz et al., 2021), outperforming the benchmarks in late nowcast periods once real activity indicators for the nowcast quarter are released. The relative improvement is particularly pronounced once the Covid-19 shock is included in the sample, where our model can detect the initial trough in GDP earlier on and nowcasts the economic recovery more precisely.

Third, we unpack intuition on the drivers of the nowcast performance via inclusion probabilities of the indicators into the model over the data release cycle. Our proposed combination of GIGG prior with sparsification is shown to read in a sparse subset of indicators throughout the data release cycle. Early nowcasts are determined by signals from survey indicators during early nowcasts, which then shifts towards signals from a few real activity indicators in later nowcasts, particularly the index of services. When including the pandemic, the model concentrates more

heavily on service-related surveys and activity indicators, as well as indicators from the housing sector, both of which reflected disruption from lock-downs and consumption shifts during the pandemic. Other models, on the other hand, continue to read dense signals from a wide range of indicators. As such, our model is better able to capture the heterogeneous nature of the Covid-19 shock.

We also provide sensitivity analyses which show that our model retains its competitive performance under a different MIDAS transformation (Forni et al., 2015). We also check nowcast performance during the pandemic when holding MIDAS coefficients fixed at their pre-pandemic levels, as advocated by Schorfheide and Song (2021). We find that this markedly decreases nowcast precision, which indicates that the flexible features of our model help in nowcasting the large shock.

Our finding of a strong nowcasting performance of a group-sparse model contributes to the general debate on the “illusion of sparsity” in macroeconomic forecasting (Giannone et al., 2021; Fava and Lopes, 2021; Kohns and Bhattacharjee, 2022). We find that grouping structure is important to account for in macro applications when analyzing sparsity patterns and that sparse models can outperform dense specifications in the context of our model, particularly in presence of a heterogeneous shock.

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

The remainder of the chapter is structured as follows. Section 3.2 presents the Trend-SV-t Bayesian MIDAS approach and the proposed flexible priors. Section 3.3 outlines the data set and setup of our nowcast application. Section 3.4 present the results on the in-sample model features, nowcast performance and variable inclusion probabilities. Section 3.5 investigates how well the proposed prior and sparsification algorithm generalise to a host of sparse and dense data-generating processes. Section 3.6 concludes. More details on the methodology and additional results can be found in Appendix B.1.

3.2 Trend-SV-t Bayesian MIDAS with flexible priors

In this section, we discuss our methodological contributions to the BMIDAS framework. Firstly, we outline the three main features of the model: cyclical MIDAS, trend and stochastic volatility components. Then, we present the group-shrinkage prior which we argue is well suited to the MIDAS estimation due to its ability to take the mixed-frequency grouping and correlation structure into account, alongside

its aggressive shrinkage which facilitates variable selection. Lastly, we detail our decision theory inspired variable selection approach that allows to communicate which high-frequency variable groups have the highest impact for the nowcasts.

3.2.1 The Model

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

The model takes the following state-space form:

$$y_t = \tau_t + \theta' Z_{t-h}^{(m)} + \sqrt{\lambda_t} e^{\frac{1}{2}(h_0 + w_h \tilde{h}_t)} \tilde{\epsilon}_t^y, \quad (3.1)$$

$$\begin{aligned} \tilde{\epsilon}_t^y &\sim N(0, 1), \quad \lambda_t \sim \mathcal{G}^{-1}(\nu/2, \nu/2) \\ \tau_t &= \tau_{t-1} + e^{\frac{1}{2}(g_0 + w_g \tilde{g}_t)} \tilde{\epsilon}_t^\tau, \quad \tilde{\epsilon}_t^\tau \sim N(0, 1) \end{aligned} \quad (3.2)$$

$$\begin{aligned} \tilde{h}_t &= \tilde{h}_{t-1} + \tilde{\epsilon}_t^h, \quad \tilde{\epsilon}_t^h \sim N(0, 1) \\ \tilde{g}_t &= \tilde{g}_{t-1} + \tilde{\epsilon}_t^g, \quad \tilde{\epsilon}_t^g \sim N(0, 1), \end{aligned} \quad (3.3)$$

where (3.1) is the observation equation of quarterly GDP growth and (3.2)-(3.3) describe the evolution of states, the latent trend and stochastic volatilities, respectively.

The cyclical component $Z_t^{(m)} = (z_{1,t}^{(m)}, \dots, z_{K,t}^{(m)})'$ is observed for each of the K high-frequency indicators at m , potentially non equidistant, intervals between $t - 1$ and t . In our application, we consider macroeconomic indicators which are observed at monthly frequency, so that $m = 3$. The parameters $\theta = (\theta'_1, \dots, \theta'_K)'$ measure the response to changes in the K macroeconomic indicators, each with L_k monthly lags¹. θ_k features $p_k + 1$ parameters that act as a linking function between these higher frequency lags and lower frequency observations. The functional form of these parameters further depend on the type of MIDAS employed (see for an overview of MIDAS methods Ghysels and Marcellino (2018)). Due to its parsimony, we use the linear Almon lag polynomial in spirit of Almon (1965) that has been recently popularised for high dimensional mixed frequency forecasting applications by Mogliani and Simoni (2021)². Parsimony is induced in linear Almon-MIDAS by assuming a $p_k \ll L_k$ polynomial process of the coefficients across high-frequency

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

² Foroni and Marcellino (2014) show that linear MIDAS methods are competitive with non-linear MIDAS weighting schemes such as the non-linear Almon and beta functions Ghysels et al. (2007), but have the advantage of being compatible with off the shelf shrinkage methods.

lags, which can be further improved with economically relevant end-point restrictions (Smith and Giles, 1976). For our application below, we assume 5 monthly lags (s.t $p_k + 1$ spans 6 months in total), a third degree polynomial ($p_k = 3$) as proposed by Mogliani and Simoni (2021) and two end-point restrictions to ensure that the weight profile peters out smoothly to 0 (see Appendix B.1 for details). Although Almon lags save degrees of freedom, the cyclical component can still suffer from multicollinearity due to the high degree of serial correlation present in the mixed frequency lags $Z_t^{(m)}$. This will be directly addressed with our shrinkage prior in section 3.2.2.

The trend component in (3.2) dictates driftless random walk behaviour which captures the low frequency changes in GDP growth. Driftless trends in GDP growth, with state variances tightly controlled by priors to enforce slow-moving change, have been shown to markedly improve nowcasts in applications to the US (Antolin-Diaz et al., 2017, 2021). This reflects the empirical finding of changes in long-run GDP growth (Kim and Nelson, 1999; McConnell and Perez-Quiros, 2000; Jurado et al., 2015) that indicate slow moving shifts in general economic conditions that are otherwise not well captured by high-frequency information.³

Finally, we allow for stochastic volatility processes \tilde{h}_t, \tilde{g}_t for GDP growth y_t and the latent trend τ_t , respectively. Volatility clustering is an empirical regularity in aggregate macroeconomic time-series data (Stock and Watson, 2007), and has been shown to significantly improve density and point forecasts in many forecasting studies (Koop and Onorante, 2019; Huber et al., 2019) and mixed frequency models (Carriero et al., 2015; Antolin-Diaz et al., 2017). The volatility processes follow driftless random walks (3.3), and are non-centred after Frühwirth-Schnatter and Wagner (2010). This exerts stronger shrinkage on the state standard deviations, w_h and w_g so as to control the variability of the state process.⁴ The more commonly employed centred SV process h_t , can be exactly recovered by noticing that $h_t = h_0 + w_h \tilde{h}_t$ (likewise for g_t). Despite the stronger shrinkage of the state variation, random walk behaviour might be inappropriate to capture the recent unprecedented size and short-livedness of the Covid-19 shock. We therefore allow the observation equation to have fat-tails via λ_t which enforces t-distributed errors with ν degrees of freedom. This discounts large contemporaneous movements in y_t and thus limits the propagation of outliers to the posteriors of the model components.

While the role of long-run trend dynamics and stochastic volatilities with fat tails for nowcasting and modelling GDP growth has recently been documented for VARs

³ A different non-stochastic approach to disciplining models to account for long-run growth features is presented in Giannone et al. (2019). Here, the authors enforce iterative forecasts to return to long-run cointegrating equilibria.

⁴ Since w_g and w_h , directly appear in the observation and state equation, which control the overall state smoothness, one can apply conjugate normal priors which exert stronger shrinkage than inverse-gamma priors. This approach has been used in many large time-varying parameter models (Huber et al., 2019; Chan, 2017b; Koop and Onorante, 2019).

and mixed-frequency dynamic factor models Lenza and Primiceri (2022); Antolin-Diaz et al. (2017, 2021), to our knowledge we are the first to incorporate time-varying trends and stochastic volatility processes into a MIDAS setting.

3.2.2 Bayesian Setup

To regularise parameter estimation variance, we make use of Bayesian priors. Regularisation is key since both the latent states as well as the multivariate MIDAS structure are highly parameterised. We first present the general prior framework that gives rise to the conditional posteriors, before detailing in 3.2.2 the form of the proposed GIGG prior and give further intuition about its behaviour.

We consider priors of the form:

$$\pi(\Upsilon) = \pi(\theta)\pi(\boldsymbol{\tau})\pi(\tilde{\mathbf{h}})\pi(\tilde{\mathbf{g}})\pi(\xi)\pi(\boldsymbol{\lambda}|\nu)\pi(\nu), \quad (3.4)$$

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

Priors for the cyclical component

We consider normal priors for the MIDAS parameters θ :

$$\pi(\theta) \sim N(0_{\sum_{k=1}^K (p_k+1)}, \Lambda_*), \quad (3.5)$$

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

$$\begin{aligned} \theta|\mathbf{y}, \bullet &\sim N(\bar{\theta}, \bar{\Lambda}_*^{-1}) \\ \bar{\Lambda}_* &= (\mathbf{Z}^{(m)'}\Lambda_t^{-1}\Lambda_h^{-1}\mathbf{Z}^{(m)} + \Lambda_*^{-1}), \quad \bar{\theta} = \bar{\Lambda}_*^{-1}(\mathbf{Z}^{(m)'}\Lambda_t^{-1}\Lambda_h^{-1}\tilde{\mathbf{y}}), \end{aligned} \quad (3.6)$$

where $\tilde{\mathbf{y}} = \mathbf{y} - \boldsymbol{\tau}$, $\Lambda_t = \text{diag}(\lambda_1, \dots, \lambda_T)$, $\Lambda_h = \text{diag}(e^{h_1}, \dots, e^{h_T})$, and $\mathbf{Z}^{(m)} = (Z_1^{(m)}, \dots, Z_T^{(m)})'$. 3.6 highlights the effects of the stochastic volatility and fat-tail imposed in the model. Both $\tilde{\mathbf{h}}$ and $\boldsymbol{\lambda}$ act to effectively discount variation in the covariance matrix as well as the fit with $\tilde{\mathbf{y}}$. The scales in Λ_t move inversely with irregular shocks in conditional GDP growth, while Λ_h account for slow moving error variance.

Due to the availability of known conditional distributions, we make use of Gibbs sampling to draw inference on the parameters (see B.1.2), where we use 5000 draws as burn-in and retain 5000 for inference. In situations where Λ_* is very high-dimensional, particularly when $\sum_{k=1}^K p_k + 1 \gg T$, we make use of the fast sampling algorithm of Bhattacharya et al. (2016) which reduces computational complexity for the regression parameter sampling step from $\mathcal{O}((\sum_{k=1}^K p_k + 1)^3)$ (Cholesky based sampling algorithms) to $\mathcal{O}(T^2 \times (\sum_{k=1}^K p_k + 1))$.⁵

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

The multivariate MIDAS framework is highly parameterised since it involves K indicators, and for each indicator k , a group of L_k lower frequency (quarterly) vectors are created corresponding to each available observation at higher (monthly) frequency. This gives two potentially interdependent dimensions relevant for efficient shrinkage. First, not all of the K groups are equally relevant for nowcasting which calls for adaptive *shrinkage across groups*. Second, the fact that consecutive higher frequency observations enter the model via the implied data transformations induces correlation *within the lag group*. Such correlation can be high for time-series applications, even with U-MIDAS sampled data (Ghysels et al., 2007), and more so when Almon lag polynomials impose a structure on the lags. This can cause mixing issues with aggressive shrinkage priors which typically seek to shrink individual covariates aggressively to zero or not at all (Pironen et al., 2020). Finally, the two dimensions can be interdependent, since the degree of correlation across lags can matter for the relative impact that the lag group as a whole has for the nowcast target.

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

$$\begin{aligned} \theta_{k,j} &\sim N(0, \vartheta^2 \gamma_k^2 \varphi_{k,j}^2), \quad \forall j \in \{0, \dots, p_k + 1\} \\ \vartheta &\sim C_+(0, 1), \quad \gamma_k^2 | a_k \sim \mathcal{G}(a_k, 1), \quad \varphi_{k,j}^2 \sim \mathcal{G}^{-1}(b_k, 1), \end{aligned} \quad (3.7)$$

where $\mathcal{G}(\bullet, \bullet)$, $\mathcal{G}^{-1}(\bullet, \bullet)$ and $C_+(\bullet, \bullet)$ refer to the Gamma, inverse-Gamma, and half Cauchy distribution with positive support, respectively⁶.

The parameters ϑ, γ_k and $\varphi_{k,j}$ govern the three-tiered shrinkage. While ϑ controls the overall level of sparsity, γ_k^2 acts as a shrinkage factor that enables pushing the impact of group k jointly close to zero, and $\varphi_{k,j}^2$ controls how correlated group member j is within k . The hyper-parameters a_k and b_k summarise our prior guess

⁵ Details of the sampling algorithm can be found in B.1.1

⁶ Boss et al. (2021) formulate the prior under the assumption of dependence with the error variance, whereas in this chapter, the prior is formulated under independence. This assumption is needed to obtain the conditional posteriors analytically.

on the relative importance of group-level shrinkage relative to shrinkage of within-group correlation (see Appendix B.1 for a visualisation of the a-priori behaviour). The lower a_k is set relative to b_k , the stronger the group-level shrinkage, and the larger the prior correlation among the group’s regression parameters. Such a choice favors group-sparse posteriors, where groups are shrunk jointly to zero, but the relevant groups feature a high degree of correlation among individual lags, instead of a heterogeneous shrinkage within group. In this, setting small a_k and relatively large b_k can be a good fit for models that use Almon transformed lags, and even in unrestricted MIDAS applications where lags are highly correlated, as suggested by Boss et al. (2021). In our application, we follow this intuition and we set $a_k = 1/T, b_k = 0.5 \forall k$.

One the other hand, in situations in which prior knowledge exists that only selected lags are important, a relatively small b_k may help the prior to shrink individual lags within a group more idiosyncratically toward zero. As a robustness check, we report in the appendix (B.2.3) nowcast cast results for different hyper-parameter choices.

The GIGG prior belongs to the general global-local prior framework, in which two heavy-tailed scale processes simultaneously shrink globally and on individual covariate level (see Polson and Scott (2010) and Polson et al. (2014)). Among these, the horseshoe prior of Carvalho et al. (2010) is particularly popular due to its excellent empirical and theoretical properties (Bhadra et al., 2019; Van Der Pas et al., 2014). Conventional global-local shrinkage priors such as the horseshoe, however, do not feature group shrinkage via γ_k^2 and can suffer from random covariate selection and bad mixing for highly correlated designs (Boss et al., 2021; Piironen et al., 2020), which is typically the case in mixed-frequency regressions. Previous BMIDAS priors such as in Carriero et al. (2015) apply shrinkage on individual level with implicit grouping via a Minnesota type prior to address this, but the Minnesota prior does not impose adaptive shrinkage across covariates k and imposes a deterministic level of penalisation that increases with the lag-length. Closest to our prior setup for MIDAS regression is Kohns and Bhattacharjee (2022) who use simple horseshoe prior regularisation. We will use this prior as a benchmark for the group-prior model.

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

⁷ This represents a departure from the group-horseshoe prior of Xu and Ghosh (2015), which does not reduce to a simple horseshoe for a group size of 1.

for group selection tasks (Piironen et al., 2017).⁸ The shrinkage properties of the prior are further elucidated in Appendix (B.1.1).

In contrast to the proposed GIGG prior, other group-prior implementations such as the group-lasso suggested in the literature (Casella et al., 2010; Xu and Ghosh, 2015), are known for over-regularising signals due to their exponential tails and under-regularising noise due to less mass on zero compared to horseshoe type priors. Xu and Ghosh (2015) remedy the latter behaviour of the group-lasso by including group-spike-and-slab variable selection. However, these priors assume a uniform level of shrinkage within group, hence, don't allow for inference on the correlation via a covariate level scale. Further, the spike-and-slab with point-mass on zero is well known to mix poorly in high dimensions and with correlated groups (Ishwaran et al., 2005; Barbieri et al., 2021; Piironen et al., 2020). For these reasons, we do not consider lasso priors, but refer to Boss et al. (2021) for further prior comparisons to the GIGG.

Priors for the latent states & fat-tails

The other priors we employ are standard. For the latent states $(\boldsymbol{\tau}, \tilde{\boldsymbol{h}}, \tilde{\boldsymbol{g}})$ we consider a joint normal prior derived using methods proposed in Chan and Jeliazkov (2009); McCausland et al. (2011) that allow representing the entire conditional state posterior as a tractable normal distribution. This increases sampling efficiency compared to Kalman filter based techniques such as Carter and Kohn (1994); Frühwirth-Schnatter (1994), and allows use of computationally efficient sparse matrix operations (see B.1.1 for a more detailed exposition). To derive the posteriors of $(\tilde{\boldsymbol{h}}, \tilde{\boldsymbol{g}})$ we use the approximate sampler of Kim et al. (1998). Since latent states are prone to overfitting in heavily parameterised models Frühwirth-Schnatter and Wagner (2010) we put normal priors on the state variances to control their variation. These exert stronger shrinkage than commonly employed inverse-Gamma priors for variance parameters (see Frühwirth-Schnatter and Wagner (2010) for more discussion on state variance priors). The conditional posteriors are further exposed in B.1.1. Lastly, for the degrees of freedom hyperparameter ν of the Inverse-Gamma distribution of λ_t we assume a relatively non-informative uniform prior which results in a non-standard distribution detailed in appendix B.1.1. To draw from that posterior, we make use of a Metropolis-within-Gibbs step.

3.2.3 Sparsification Step for the GIGG Prior

With continuous shrinkage priors such as the GIGG the posteriors of lag groups remain non-zero with probability one (Hahn and Carvalho, 2015). This hampers the understanding of which indicators impact the cyclical component and thus the

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

interpretability of the nowcasts. Thresholding those lag groups with little effect on y_t to exactly zero, i.e. essentially eliminating them from the model, makes it easier to interpret and communicate model outcomes and allows to draw inference about model uncertainty which can be pervasive in macroeconomic applications (Giannone et al., 2021; Huber et al., 2019; Cross et al., 2020; Kohns and Bhattacharjee, 2022). Recently, Mogliani and Simoni (2021) extend the adaptive group-lasso prior applied to Almon-lag MIDAS regressions to spike-and-slab variable selection for that purpose. Instead, we propose the use of a sparsification algorithm that is motivated by the perspective of a Bayesian decision maker who seeks the smallest subset of groups that best summarise the forecasts of the model (3.1)-(3.3), inspired by Hahn and Carvalho (2015); Woo and Owen (2019).⁹ This can be achieved by minimising a utility function over the Euclidean distance between a linear model that penalises group-size akin to (Zou, 2006) and our model’s nowcasts:

$$\mathcal{L}(\tilde{\mathbf{Y}}, \psi) = \frac{1}{2} \|\mathbf{Z}^{(m)}\psi - \tilde{\mathbf{Y}}\|_2^2 + \sum_{k=1}^K \zeta_k \|\psi_k\|_2, \quad (3.8)$$

where $\tilde{\mathbf{Y}}$ refers to a realisation from the posterior predictive distribution $p(\tilde{\mathbf{Y}}|\mathbf{y}) = \int p(\tilde{\mathbf{Y}}|\mathbf{y}, \mathbf{Z}^{(m)}, \theta, \bullet) p(\theta|\mathbf{y}, \mathbf{Z}^{(m)}, \bullet) d\theta$, and $\|\bullet\|_p$ refers to the ℓ_p -norm.¹⁰ Similar to the logic of adaptive group-lasso (Wang and Leng, 2008), the penalisation term induces non-differentiability at zero, which creates a soft-thresholding effect between $[-\zeta_k, \zeta_k]$, thereby forcing the coefficients on all group members to zero. The Bayes optimal solution for ψ is obtained by integrating out the posterior uncertainty from the predictive distribution, as well as in the parameters θ (Lindley, 1968) (see Appendix B.1.3).

In the following, we first show the analytical solution we derive for (3.8), and then discuss the assumptions needed to derive it. For a full derivation, see appendix B.1.3. The sparsified estimate $\psi_k^{*(s)}$ for each Gibbs-sampling step $s = 1, \dots, S$, is given by:

$$\psi_k^{*(s)} = \left(\|\theta_k^{(s)}\|_2 - \zeta_k^{(s)} \right)_+ \frac{\theta_k^{(s)}}{\|\theta_k^{(s)}\|_2}, \quad (3.9)$$

where $(x)_+ = \max(x, 0)$. (3.9) implies that when $\theta_k^{(s)}$ are close to $\mathbf{0}$, then $\psi_k^{*(s)} = \mathbf{0}$, whereas, when $\theta_k^{(s)}$ are large, then $\psi_k^{*(s)} = (1 - \frac{\zeta_k^{(s)}}{\|\theta_k^{(s)}\|_2})\theta_k^{(s)}$, in which case the first term will be very close to 1, thus imposing close to no further shrinkage.

Two assumptions are needed to derive (3.9). Firstly, it requires orthonormalisation of the data for each k such that $T^{-1}\tilde{\mathbf{Z}}_k^{(m)}, \tilde{\mathbf{Z}}_k^{(m)} = \mathbf{I}$. This serves to simplify

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

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

the solution, and as shown in (Simon and Tibshirani, 2012), not orthonormalising groups ignores the cross-correlation of group members in k such that the algorithm implicitly prefers to not threshold groups whose covariance is large, and also ignores that $\mathbf{Z}_k^{(m)}$ might be on different scales.¹¹ Secondly, we make use of the work by Ray and Bhattacharya (2018); Chakraborty et al. (2020) who show that, when setting $\zeta_k^{(s)} = \frac{1}{\theta_k^{(s)}}$, iterative solution methods such as the coordinate descent (Friedman et al., 2010), converge already after the first cycle. This gives us the analytical solution 3.9.

The relative frequency of high-frequency lag-group k selected in $\psi^{*(s)}$ over all Gibbs draws will be used to report inclusion probabilities that inform on the relative impact of an indicator for nowcasting. See Woody et al. (2021) and Chakraborty et al. (2020) for formal justification of model selection uncertainty and the asymptotic risk properties for similar models, respectively.

3.3 Data Set and Nowcast Setup

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

3.3.1 Data set

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

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

exhibited a boom that was not in line with the weakness of the real economy. The series are transformed to be approximately stationary prior to estimation.¹²

We consider the sample period from 1999Q1 to 2021Q3. The start of the sample is pinned down by data availability since many of the monthly indicators are not available for earlier years.¹³ We use final vintages of the data, downloaded in December 2021. While revisions of UK monthly real activity indicators and quarterly GDP can be substantial, accounting for them accurately in the nowcast set-up also poses computational and methodological challenges.¹⁴ Since our focus here lies in understanding the proposed model, we use pseudo real time data as outlined below, and we leave an account for revisions for future research.

3.3.2 Nowcast Exercise

Macroeconomic data are published asynchronously at different points in time and with delays ranging from various weeks (survey data) to up to various months (labour market data) after the reference month. To simulate the information set available to the nowcaster over the data release cycle, we follow a stylised pseudo real-time data release calendar (see Table 3.1).

Table 3.1: Stylised pseudo real-time data release calendar.

Nowcast	Quarter	Days to GDP	Month	Timing within month	Release	Publication Lag
1		135	1	1st of month	PMIs	m-1
2		125	1	End of 2nd week	IoP, IoS, Ex, Im	m-2
3		120	1	3rd week	Labour market data	m-2
4		115	1	3rd Friday of month	Mortgage & Visa	m-1
5		110	1	End of 3rd week	CBI & 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	CBI & 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	CBI & GfK	m
17		45	1	1st of month	PMIs	m-1
18	Subsequent	35	1	End of 2nd week	IoP, IoS, Ex, Im	m-2
19	quarter	30	1	3rd week	Labour market data	m-2
20	(backcast)	25	1	3rd Friday of month	Mortgage & Visa	m-1

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

¹² See Table B.1 in the appendix for an overview of the data and their respective transformations.

¹³ Some of the series have missing values at the beginning of the sample period. We interpolate these based on a principal component (PCA) model that accounts for missing information via the alternating least square algorithm. Alternatively, we also employed the commonly used EM algorithm (Bańbura and Modugno, 2014) for interpolation, and we found that there is little difference in the sample under investigation.

¹⁴ See Anesti et al. (2018) for an analysis of UK data on the forecastability of different vintages and how to incorporate that information for nowcast updates.

As is common in the MIDAS nowcasting literature, we start the nowcast exercise for each quarter anew, where we start predicting with all available information on the first of the month prior to the reference quarter (forecasts), and simulate incoming data throughout the reference quarter (nowcasts), until GDP is finally observed six weeks after the reference quarter (backcasts).¹⁵ For each new data release over the data release cycle, we generate nowcasts from the predictive distribution $p(y_{t+1}|\Omega_T^v)$, where $(v = 1, \dots, 20)$ refers to the nowcast periods within the stylised data release calendar (Table 3.1) and Ω_T^v represents the real-time information set that expands with each data release. Since the MIDAS framework belongs to the class of reduced system mixed frequency models (Bańbura et al., 2013), each information set Ω_T^v results in a different model (depending on which data are observable to the nowcaster).¹⁶ To draw samples from the predictive distribution, we integrate over all parameter uncertainties which is easily implemented via Monte Carlo integration (Cogley et al., 2005).

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

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

$$\text{RMSFE} = \sqrt{\frac{1}{T_{end}} \sum_{t=1}^{T_{end}} (y_{T+t} - \hat{y}_{T+t|\Omega_{T+t-1}^v}^v)^2}, \quad (3.10)$$

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

$$\text{CRPS} = \frac{1}{T_{end}} \sum_{t=1}^{T_{end}} \frac{1}{2} \left| y_{T+t} - \hat{y}_{T+t|\Omega_{T+t-1}^v}^v \right| - \frac{1}{2} \left| y_{T+t|\Omega_{T+t-1}^{v,A}} - y_{T+t|\Omega_{T+t-1}^{v,B}} \right|. \quad (3.11)$$

Note that in (3.11), $y_{T+j|\Omega_{T+j-1}^{v,A,B}}$ are independently drawn from the posterior predictive density $p(y_{T+1}^v|\Omega_{T+j-1}^v|y_T)$. The CRPS belongs to the class of strictly proper scoring rules (Gneiting and Raftery, 2007), and can be thought of as the probabilistic generalisation of the mean-absolute-forecast-error. To facilitate the discussion

¹⁵ We refer to the first GDP publication, available about 40 days after the reference quarter. We abstain from accounting for a less accurate preliminary GDP estimate that was available 25 days after the reference quarter prior to Sir Charles Bean’s 2018 review of UK economic statistics (Scruton et al., 2018).

¹⁶ This is in contrast to full system mixed frequency methods (Bańbura et al., 2013) which interpolate missing low- and high-frequency information via filtering methods.

below, the objective in terms of nowcasting precision is to minimise both evaluation metrics.

3.4 Empirical Results

In section 3.4.1, we first develop intuition on the trend and cyclical components in UK GDP growth by focusing on posterior estimates of the various model components (3.1)-(3.3). Section 3.4.2, evaluates nowcast performance of the Trend-SV-t-BMIDAS model against a simpler BMIDAS without additional model features, and then against commonly employed benchmark models. Section 3.4.3 then unpacks the signals behind the nowcast estimates via inclusion probabilities derived by the decision theoretic variable selection method.

3.4.1 Analysing UK GDP Growth via The Trend-SV-t-BMIDAS

Figure 3.1 shows the posterior estimates of the cyclical (blue) and trend (orange) components (upper panels, separating the pre-pandemic period and the Covid-19 period) as well as the stochastic volatility components of GDP growth (lower left panel) and the trend equation (lower right panel) from the Trend-SV-t-BMIDAS model with GIGG prior. The cyclical component captures high frequency movements in GDP growth. It tracks the quarter-to-quarter movements in GDP growth (black dashed-dotted lines) well, including over the Covid-19 pandemic where the cyclical component captures the bulk of the 20% drop in GDP growth and most of the recovery. On the other hand, the trend captures low frequency changes in GDP growth and can be interpreted as a time-varying long-run growth estimate. We observe a gradual slowdown in UK GDP growth since the early 2000s, with an additional temporary decrease in the trend during the Great Financial Crisis (GFC). Throughout the Covid-19 pandemic, the trend remains largely unchanged, hence the model interprets the extreme movements in GDP growth seen during the pandemic as transitory in nature.

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

The fat-tailedness of error distributions in the stochastic volatility process proves

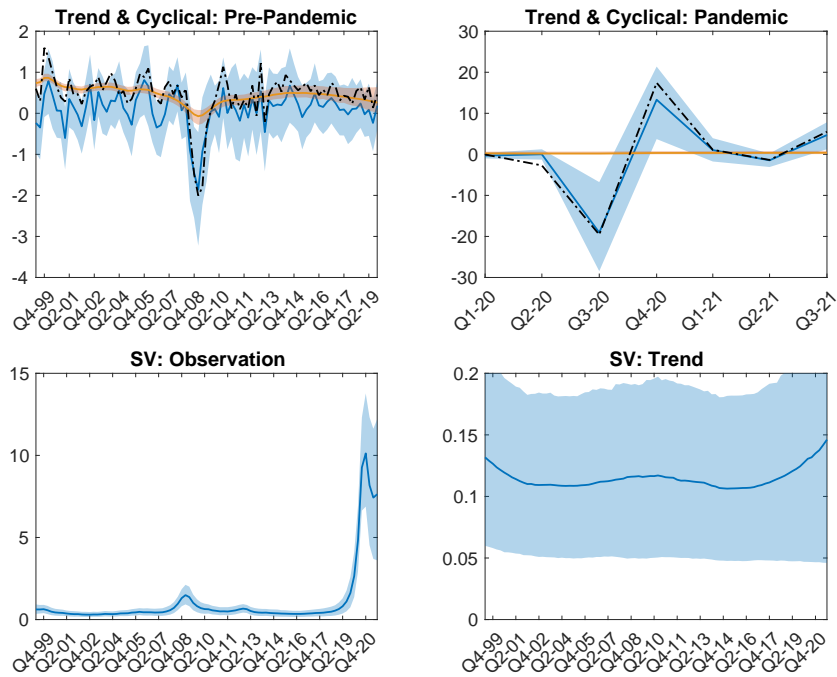


Figure 3.1: Posterior estimates for trend, cyclical component and stochastic volatilities.

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

to be an inherent model feature, as suggested by the posterior distribution of the degrees of freedom parameter showing large mass around small values (see Figure B.2 in the appendix). But even prior to the pandemic, the inclusion of fat-tailed error variance is important to separate the slow-moving trend from cyclical movement. To elucidate this point, Figure 3.2 shows the posterior trend and cyclical component over the period until 2019Q4, from the baseline model that includes SV and t-distributed errors (blue lines and shaded areas), the model with SV but no t-distr. errors (orange), as well as for constant variance only (green).¹⁷ The model with SV shows the least intuitive results since it assigns most of the cyclical variation to the trend component, whereas the cycle remains very stable throughout the sample including during the GFC. Adding SV without t-distr. errors in a sample that contains both quieter periods and larger shocks such as the GFC likely results in overly discounting variation available to the cyclical component and instead overfitting the trend. On the other hand, the baseline model achieves a sensible trend-cycle decomposition since it is able to separate large transitory shocks, that are handled by the fat tails, from smaller but more pervasive ones that are captured by

¹⁷ The posterior estimates are based on the information set at the first nowcast period of 2019Q4. As section 3.4.2 will show, early nowcasts over the data release cycle are those benefiting the most from the inclusion of a time-varying trend with SV-t errors during the pre-pandemic period.

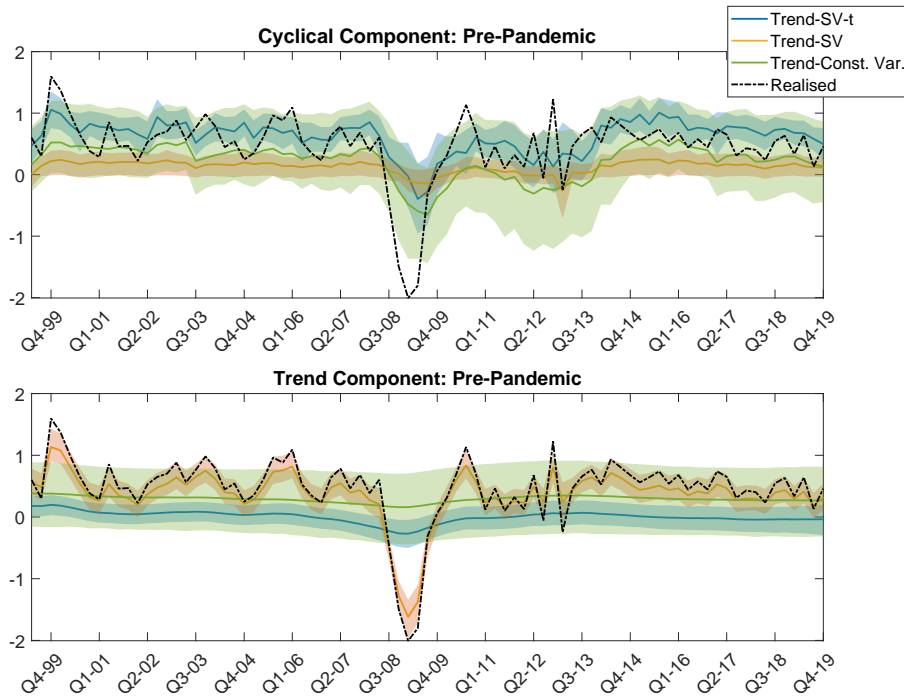


Figure 3.2: Posterior trend estimates from different BMIDAS specifications, pre-pandemic.

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

the SVs in GDP growth and its trend. Both the cyclical and trend components are estimated with high precision in the baseline model. By contrast, in the constant variance model credible intervals are much larger, since the uncertainty of shocks permeates fully into the cyclical and trend components. While this model does assign short-term fluctuations to the cyclical component, this goes at the expense of an almost unresponsive trend.

3.4.2 Nowcast Evaluation and The Role of Model Features

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

Role of model features for nowcasting

Figure 3.3 assesses the role of model components by comparing the nowcast performance of the baseline model with versions where the new features, i.e. the time-varying trend, the stochastic volatilities, and their t-distributed errors, are shut off one by one or jointly. The figure shows root mean square forecast errors (RMSFE,

upper panel) and continuous rank probability scores (CRPS, lower panel) over the data release cycle (days ahead of GDP release) on the x-axis.

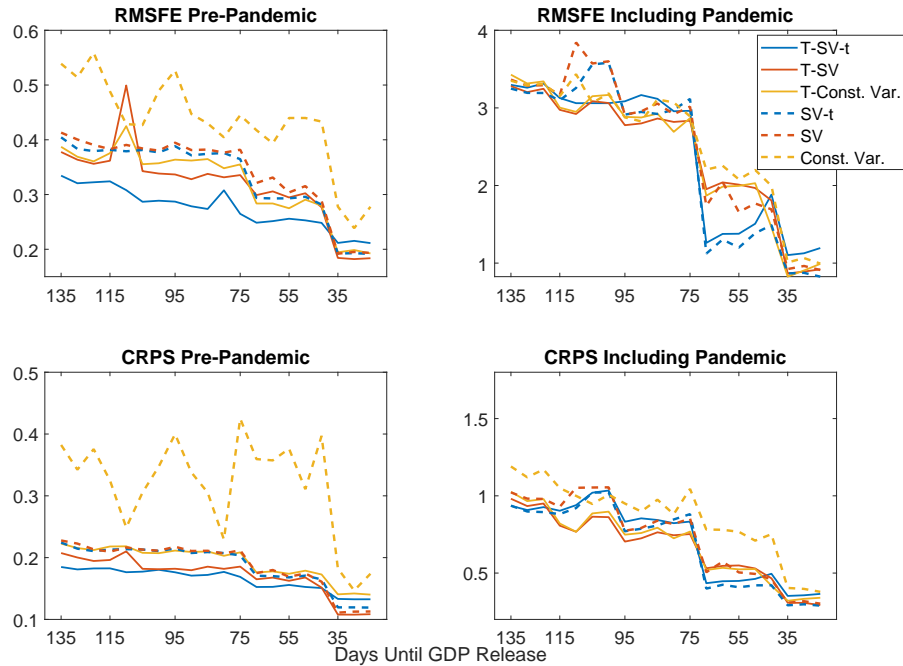


Figure 3.3: Nowcast performance for different BMIDAS model components.

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

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

that release. The model with constant variance and no trend (yellow dashed lines) clearly performs worst, which underlines the importance of incorporating at least one of the proposed features into BMIDAS models—even over samples that exclude the Covid-19 pandemic.

When including the Covid-19 pandemic (right panel), not surprisingly, nowcast errors are much higher for all models, particularly early on in the data release cycle, where the episode of the initial downturn generated by lockdowns and the health crisis inflates nowcast errors. Differences across model variants are relatively small over most nowcast periods. Nowcasts gradually improve in the early parts of the data release cycle until the release of PMIs for the second month of the reference quarter (75 days prior to GDP), followed by a strong improvement in performance with the release of “hard” indicators pertaining to the first month (65 days prior to GDP). The nowcast performance of all models strongly improves at this point, but more so for models with stochastic volatility and t-distr. errors with and without trend (blue lines). Whether the time-varying trend is added or not makes less of a difference once the pandemic period is included, in line with the findings from Figure 3.2 that models that do not account for outliers struggle to identify the Covid-19 pandemic related downturn as temporary and instead over-fit the trend. Finally, the simple model without trend and with constant variance fares comparatively well early on in the data release cycle, but then loses out against the other models, particularly in terms of density nowcasts. Such a model is not able to capture the large shift in the data neither via increased uncertainty around nowcasts nor through trend shifts.

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

Nowcast evaluation against benchmark models

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

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

- **MF-DFM**: Similar to Antolin-Diaz et al. (2021), this dynamic factor model also includes Trend-SV-t components, but it captures the cyclical information via a dynamic factor that exploits co-movement across the K indicators contemporaneously and at up to two lags. The priors on the latent state components are similar to the ones outlined in section 3.2.2.
- **Trend-SV-t-HS**: model 3.1-3.3 with the horseshoe prior, thus another flexible shrinkage prior, but without group-shrinkage. This represents a benchmark in terms of the prior for the Trend-SV-t BMIDAS.¹⁸

Figure 3.4 plots RMSFE and CRPS over the data release cycle, and Table 3.2 shows the evaluation metrics of the proposed model and benchmarks relative to the AR model, on average across all nowcast periods and for selected periods that correspond to releases of PMIs and real activity indicators. Stars indicate significantly different point and density nowcasts as per Diebold et al. (1998). Three main findings emerge. First, similarly to many other nowcast studies (e.g. Foroni et al. (2015); Carriero et al. (2015, 2019)), models that exploit monthly macroeconomic indicators outperform the autoregressive model, with relative RMSFE and CRPS against the AR(2) lying clearly and significantly below 1 for most models throughout the nowcast cycle. The relative advantage against the AR(2) model becomes even more pronounced when including the pandemic quarters into the evaluation, albeit the differences become insignificant for individual nowcast quarters due to the large uncertainty during that shock.

Table 3.2: Relative Evaluation Metrics

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

Notes: The table shows the average RMSFE and CRPS values for the AR model in the first row of each panel across all 20 nowcast periods (“Average”), and for selected nowcast periods (6,13,18). RMSFE and CRPS values for the other models are in relative terms to the AR model and stars indicate significance as per the Diebold-Mariano test Diebold et al. (1998) (* = 10% significance, **=5% significance, ***=1% significance)

Second, the proposed Trend-SV-t model with GIGG prior is among the strongest models prior to the pandemic and outperforms all other models once pandemic quar-

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

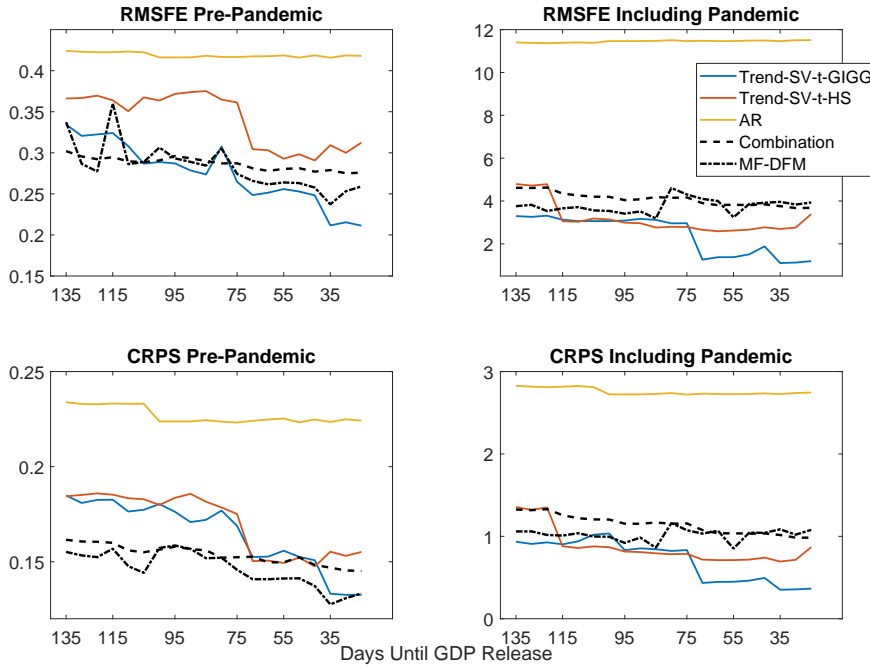


Figure 3.4: Nowcast performance of Trend-SV-t-BMIDAS with GIGG compared to benchmarks.

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

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

Third, regarding the role of the shrinkage prior, exerting group-level shrinkage and taking into account the high-frequency correlation structure via the GIGG prior (blue line) has preferable performance compared to the horseshoe (HS) prior (yellow

line). Prior to the pandemic, the model with GIGG prior has lower point nowcast errors throughout and almost always better density fit. When including the pandemic quarters, it is strongest for the first nowcast periods and again starting from 75 days ahead of GDP publication. As we illustrate via variable inclusion probabilities in section 3.4.3, this nowcast gain is achieved because the GIGG prior shrinks the information set towards a more sparse selection of indicators, and because it shifts its signal extractions towards indicators related to the service sector when the pandemic hits.

Nowcasts during the Covid-19 pandemic

At which point during the pandemic and the data release cycle does the proposed Trend-SV-t BMIDAS model with GIGG prior achieve its forecast gains? Figure 3.5 visualises the nowcasts over time for the proposed model compared to the sets of models discussed in section 3.4.2, for selected nowcast periods. Pre-pandemic, nowcasts are fairly close to each other. The Trend-SV-t BMIDAS nowcasts show somewhat more volatility, which can explain their slightly weaker performance compared to Combined MIDAS and DFM. For the Covid-19 pandemic quarters, unsurprisingly, all models miss the large unprecedented trough early in the data release cycle. However, the Trend-SV-t BMIDAS with GIGG prior is the quickest to update nowcasts to the trough and subsequent rebound in activity, and its nowcasts in later nowcast periods are closest to the actual realisation. In early nowcast periods, it is the only model to indicate the large rebound in activity for Q3-2020. And once the real activity indicators for the first reference month have been published, it shows the largest downward adjustment for Q2-2020. The other models are less responsive, or belatedly nowcast a trough without capturing the recovery.¹⁹

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

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

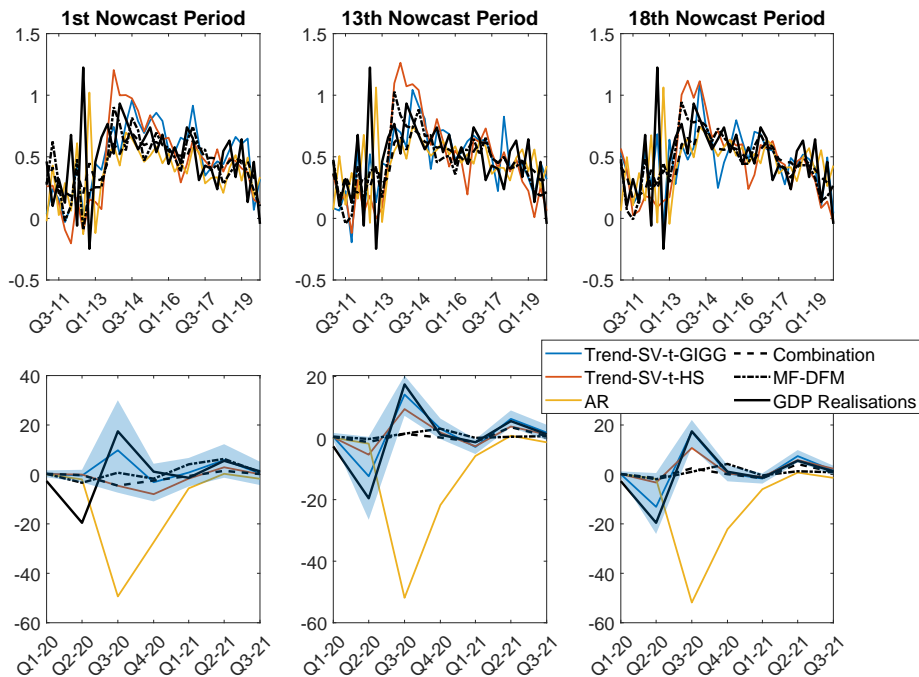


Figure 3.5: Posterior mean and density nowcasts, selected nowcast periods.

Notes: Upper panel shows nowcasts over time for the period 2011Q1 to 2019Q4 (x-axis) at different points in the data release cycle (columns), lower panel shows nowcasts over the pandemic quarters on the x-axis. Models are the same as in Figure 3.4. Shaded areas refer to Trend-SV-t-GIGG model and show 95% credible interval. Black solid lines show quarterly GDP growth realisations.

3.4.3 Interpreting Nowcasts via Variable Inclusion Probabilities

The group-variable selection achieved via the sparsification algorithm (3.9) equips us with an intuitive way to communicate nowcast signals within a Bayesian, decision theoretically based setting. Variable selection is communicated as the inclusion probability of high-frequency lag group k , i.e. a macroeconomic indicator and its lags, into the linear model (3.8) over posterior samples. The higher the inclusion probability of lag group k , the larger its impact onto the predictions of the model. Inclusion probabilities turn out rather stable over the sample period before the Covid-19 pandemic, so that we focus on averages over sub-samples.

Figure 3.6 presents heatmaps for inclusion probabilities for each indicator (x-axis) over nowcast periods (y-axis). The lower sub-plots in each panel shows corresponding results using the horseshoe prior without group-shrinkage. Pre-pandemic (panel a), the model with GIGG prior (upper sub-plot) selects a sparse specification, as indicated by only a few indicators shaded in red. The model uses signals from one to three indicators at a time, shifting to signals from other indicators as they get released. Early in the nowcast cycle the model mostly exploits a few survey variables: Manufacturing and Construction PMIs for very early nowcasts and then GfK consumer confidence for the first month of the reference quarter, once it gets

released in nowcast period 5. Labour market data also plays a role when released in period 9. However, once the Once, however, “hard” real economic information get published 65 days prior to GDP release (period 13), the model loads almost exclusively on the index of services. This agrees with earlier nowcast studies who find that survey indicators provide the main early signals for quarterly GDP, but that incoming hard information becomes more important for nowcasts later in the data release cycle (Bańbura et al., 2013; Carriero et al., 2015; Anesti et al., 2017).

When the pandemic is included into the sample (panel b), both prior specifications make the model read a wider range of signals as indicated by the overall darker colours. Nonetheless, the model with GIGG prior remains much more sparse, with a clear pattern of exploiting different signals over the data release cycle. It keeps exploiting survey indicators in the early part of the sample, but interestingly the most informative indicator now becomes the Service PMI, followed by Construction PMI. Over nowcast periods 5 to 12, signal from mortgage lending are relevant too. Little focus is put on the GFK and labour market data. Once hard economic data become available, very strong signals are read from the index of services, and additional signals ahead of GDP release come from the Service PMI and index of production. In this, the model efficiently exploits that during the pandemic most of the disruption to the economy was stemming from lockdowns affecting the service sector as well as initially a shut-down of the housing and construction sectors, whereas consumer confidence and manufacturing remained much less affected and labour market data were distorted by the furlough scheme. This likely helps the model with GIGG prior to capture the recovery from the Covid-19 induced trough early in the data release cycle and to update its nowcast of the initial trough in GDP earlier compared to other models, as discussed in section 3.4.2.

On the other hand, the model with horseshoe prior (lower sub-plots) shows dense inclusion pattern both pre-pandemic and particularly once the pandemic is include. It draws on signals from surveys, real, labour and personal finance indicators in a diffused way over the data release cycle. The denser signal extraction with the horseshoe prior could stem from increased cross-sectional correlation in the face of large macroeconomic shocks, documented in studies such as Rockova and McAlinn (2021); McAlinn et al. (2018); Frühwirth-Schnatter and Lopes (2018). However, given that the Covid-19 shock affected specific sectors more than others, a dense solution can represent a disadvantage. And the heavy reliance of the model with horseshoe prior on surveys even when hard economic information is available can explain the relatively weak performance of the model discussed in section 3.4.2.

Overall, this illustrates how group shrinkage helps exploiting signals from highly correlated macroeconomic data. In this, our findings provide new impetus to the debate on the “illusion of sparsity”, where models with a dense cyclical component have been found to forecast better in applications with macroeconomic data than

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

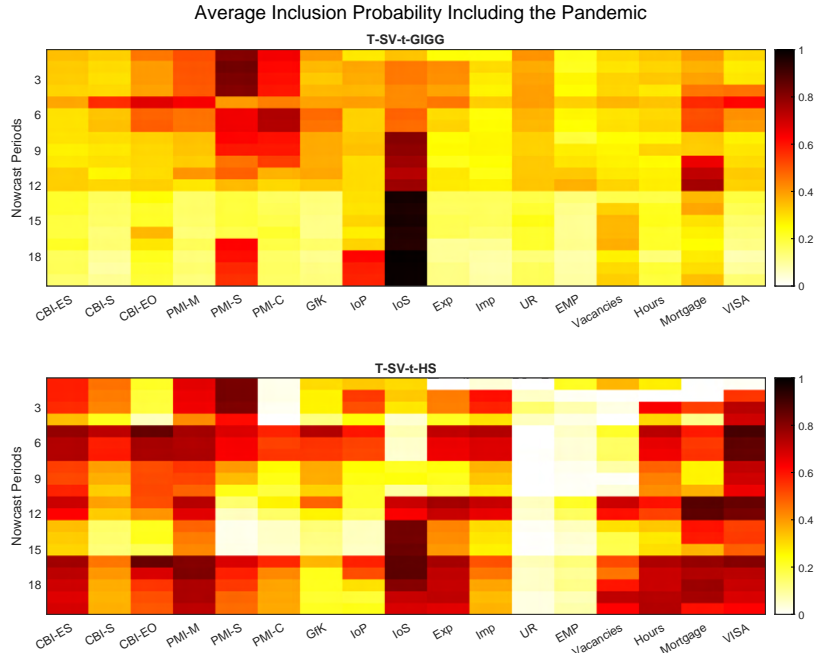
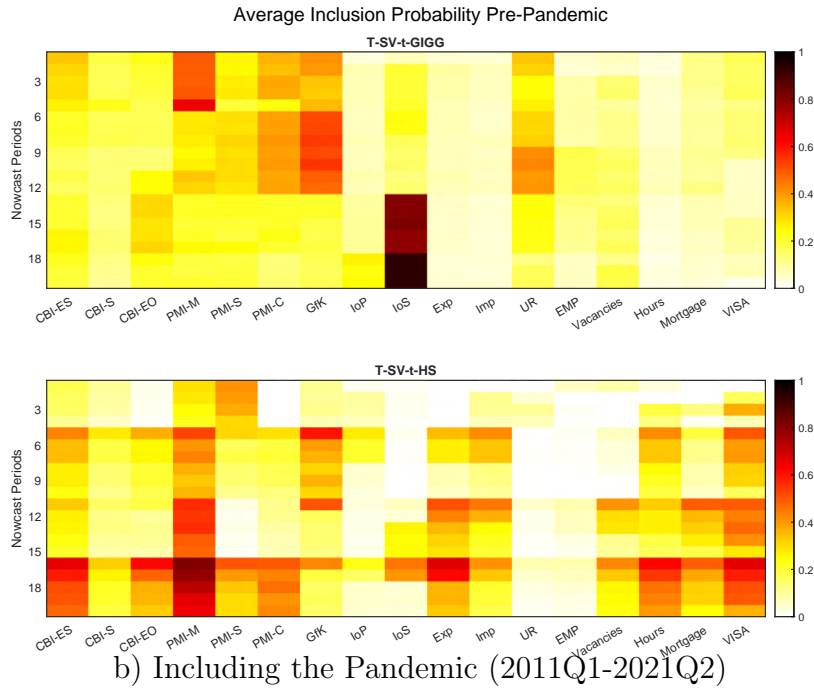


Figure 3.6: Average posterior inclusion probabilities over high frequency lags for each indicator.

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

models that prefer sparse model solutions (Giannone et al., 2021; Fava and Lopes, 2021; Kohns and Bhattacharjee, 2022; Cross et al., 2021). We find, however, that denser specifications, as the horseshoe prior model, or the MF-DFM, do not necessarily forecast better than the group-sparse model with GIGG prior, particularly in presence of the Covid-19 shock affecting the economy heterogeneously. Similar results for forecasting applications have been found for comparable aggressive shrinkage priors to the horseshoe in Fava and Lopes (2021) who show elevated variable selection uncertainty with strongly correlated data (Pironen et al., 2020).

3.4.4 Sensitivity Analyses

We conduct two sensitivity analyses, one with respect to the stability of nowcast performance of our preferred model when holding coefficient fixed over the Covid-19 pandemic, and the other using an unrestricted MIDAS specification instead of Almon lag polynomials. Nowcast evaluation results in comparison to the baseline specification are presented in Figure (3.7).

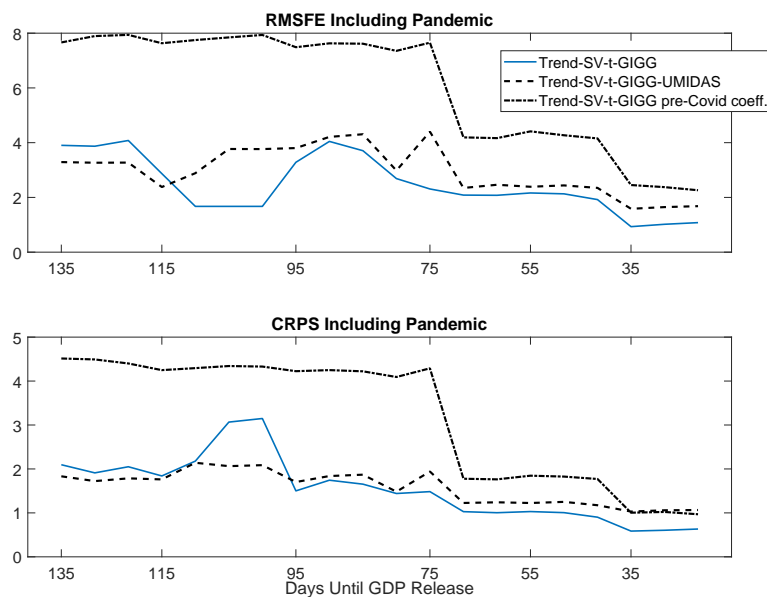


Figure 3.7: Nowcast performance of alternative specifications of the Trend-SV-t-BMIDAS.

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

Holding model coefficients fixed over the Covid-19 pandemic is motivated by Schorfheide and Song (2021) who show that MF-VAR models without flexible error components can achieve similar nowcast performance (after the trough) by simply omitting the problematic first two quarters of 2020 in the estimation sample, which is similar to using scale processes that inversely move with periods of large volatility, thereby down-weighting their effect. We nowcast with our model starting Q1-2020

with parameter coefficients based on the sample ending with Q4-2019. If the GDP dynamics during the Covid-19 shock truly were only outliers in the UK, the nowcast performance for Q3-2020 to Q2-2021 should be similar to those with the original Trend-SV-t-GIGG model. However, we find that using pre-pandemic coefficients, the model performs significantly worse both in point as well as density nowcasts up until about 40 days until GDP release, and then shows similar performance. As discussed in section 3.4.3, with onset of the pandemic, the baseline model with GIGG prior shifts variable selection towards indicators reflecting sectors that were hit strongly by the shock. These dynamics are not picked up when the Covid-19 period is omitted from the estimation.

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

3.5 Simulation Study

In the nowcasting exercise using the GIGG prior, we find evidence for a sparse set of groups that drive the cyclical component. Since this is in opposition to the existing canon in the macro forecasting literature (Giannone et al., 2021; Cross et al., 2020) which finds that dense components usually drive forecasts of aggregate economic time-series, we investigate in this section parameter estimation, variable selection and forecasting precision via a simulation study. In particular, we are interested in how our model’s performance generalises to a variety of different MIDAS data-generating-processes (DGPs) where we change correlation structures, degrees of sparsity and MIDAS lag profiles. And to abstract from any estimation issues stemming from time stochastic components, we consider in this section only DGPs without Trend and SV-t components. This also serves to facilitate comparability to similar simulation studies such as in Mogliani and Simoni (2021). We generate data from the following model:

$$y_t = \beta_0 + \sum_{k=1}^K \beta_k \sum_{j=0}^{p_k+1} \theta_j z_{j,k,t-h}^{(m)} + \epsilon_t^y, \quad \epsilon_t^y \sim N(0, \sigma^2) \quad (3.12)$$

$$x_{k,t-h}^{(m)} = \mu + \rho x_{k,t-h-1}^{(m)} + \epsilon_{k,t}, \quad k = 1, \dots, K,$$

where $x_{k,t}^{(m)}$ are the un-transformed high-frequency data ($X_t^{(m)} = (x_{1,t}^{(m)}, \dots, x_{K,t}^{(m)})$), and $z_{k,t}^{(m)}$ are the MIDAS transformed high-frequency lags. This notation distinguishes between the MIDAS lag profile, modelled via normalised weights θ_j , and the impact it has on y_t , β_k . By including zero entries in β_k , we are able to control the group-wise degree of sparsity. High-frequency data are considered in monthly frequency, $m = 3$, and we consider $L_k = 24$ monthly lags, to make the simulation study comparable to Mogliani and Simoni (2021)²⁰. The correlation structure of the high-frequency data, $X_t^{(m)}$, are determined by $\epsilon_t^K = (\epsilon_{1,t}, \dots, \epsilon_{K,t}) \sim N(0, \Sigma_K)$, where Σ_K has elements $\sigma_k^{|k-k'|}$.

We vary the DGP along 4 different dimensions: 1) the number of groups, K , 2) the sparsity in β , 3) cross-sectional high-frequency correlation σ_k , and 4) the MIDAS lag profile in θ . We consider relatively small to large dimensions, $K = \{30, 50, 100\}$. For β , we generate from a sparse regime $\beta = (0, b_1, b_2, 0, b_3, b_4, 0, 0, b_5, \mathbf{0})$, and a dense regime where we populate randomly 70% of the entries with non-zero coefficients. For both sparsity regimes, we generate $b_k \sim N(0, 1)$ such that both small and large coefficients are present in the DGPs. To test the prior in both low and high correlation data environments, we set $\sigma_k \in \{0.5, 0.9\}$. Finally, for the weight profiles θ , we follow Andreou et al. (2010) and Mogliani and Simoni (2021) by considering normalised weights which follow a steep, moderately steep and flat decay across the 24 high frequency lags, visualised in Figure 3.8.²¹ The profiles also serve to test how well the prior adapts to varying degrees of within-group sparsity. The steep profile sets coefficients for most lags to 0, while the flat profile's coefficients are entirely non-zero. We generate 100 simulated data sets for each DGP variant.

We specify the GIGG prior, as in the empirical application with the hyper-parameters set to $a_k = 1/T, b_k = 0.5, \forall k$, and apply the sparsification algorithm in 3.9 at each Gibbs draw. To put the performance of the GIGG prior into perspective, we consider the non-grouped horseshoe prior with group-sparsification as in the empirical application, as well as the recently proposed group prior of Mogliani and Simoni (2021), the group-adaptive-lasso prior with spike-and-slab sparsification (GAL-SS). The spike-and-slab paradigm used in the GAL-SS model is considered the canonical method for Bayesian variable selection and has been recently re-popularised in the macroeconomic forecasting literature after Giannone et al.

²⁰ We tested the simulations with $L_k = 6$ as in the empirical application. The results are comparable and available upon request

²¹ The weights are generated from a non-linear Almon function. See for more details Andreou et al. (2010).

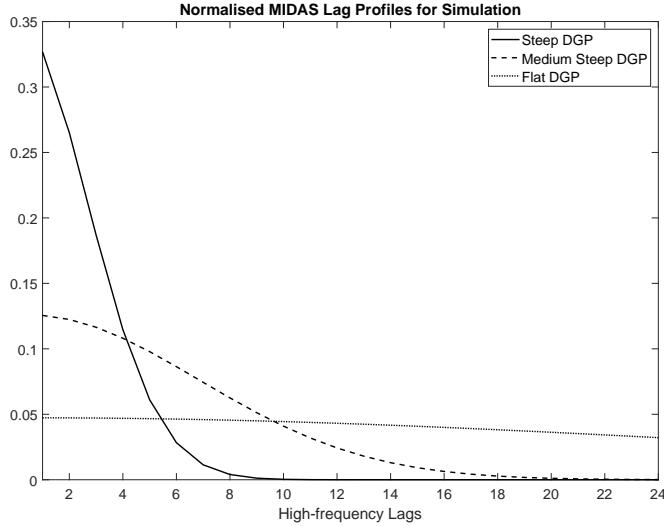


Figure 3.8: MIDAS lag profiles for a steeply, medium steeply, and flatly decaying high-frequency lag profiles over 24 for lags. These are generated from a normalised non-linear Almon function as in Andreou et al. (2010).

(2021) who have shown that the priors’ flexibility to nest dense and sparse model posteriors, outperforms sparsity favouring modelling techniques for model selection and forecasting in aggregate macroeconomic data. Mogliani and Simoni (2021) show that the GAL-SS compares favourably to competing lasso and group spike-and-slab shrinkage priors. The GAL-SS, similar to the GIGG prior, has an adaptive group-level shrinkage scale, but, importantly, does not feature an individual covariate level shrinkage scale to model the degree of correlation between the mixed frequency parameters (see Mogliani and Simoni (2021) for further details). Further, while the GAL-SS conducts variable selection via group-spike-and-slab²², our approach does variable selection via a decision theoretic approach. Priors on the error variance, σ , for all models are assumed to be uninformative.

3.5.1 Evaluation

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

$$\frac{1}{M} \sqrt{\|\hat{\theta} - \theta\|_2^2}, \quad (3.13)$$

where M is the number of simulations, set to 100, and $\hat{\theta}$ is the posterior mean based on the retained MCMC draws. To support the information about zero (i.e. “inactive”) vis-a-vis non-zero (“active”) components of each DGP, we report RMSE for each set also separately as $RMSE(\mathcal{A}^C)$ and $RMSE(A)$, respectively. To gauge

²² This can be shown to be linked to model marginal likelihoods, which, however, can be very sensitive to the hyperparameters of the prior (Barbieri and Berger, 2004; O’Hara and Sillanpää, 2009).

nowcasting performance, we generate predictions for the last observation of each sample, and treat the information set as fully available. These are evaluated via the RMSFE based on the the mean of the posterior predictive distribution, as as the CRPS, following 3.11. Predictions are generated, as in the empirical application via Monte Carlo integration.

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

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad (3.14)$$

$$TPR = \frac{TP}{TP + FN} \quad (3.15)$$

$$FPR = \frac{FP}{FP + TN} \quad (3.16)$$

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

3.5.2 Simulation Results

Tables 3.3, 3.4, and B.2 in the appendix give the average evaluation metrics over the Monte Carlo simulations.²³ Table 3.3 focuses on the sparse and table 3.4 on the dense DGPs. We show the results of the horseshoe prior (bold) in levels as the benchmark, to which the grouped priors are shown in relative terms. To facilitate the discussion, values lower than 1 for RMSE (incl. active and inactive), FPR, RMSFE and CRPS indicate out-performance of the group prior relative to the horseshoe, while values higher than 1 for MCC and the TPR entail out-performance. And to summarise the results, we plot in figure 3.9 the average relative evaluation metrics which is pooled across all but the coefficient dimensions. We will first review the general findings based on the horseshoe prior and then compare that to the performance of the GIGG and GAL-SS.

²³ We leave the medium steep results to the appendix as they are generally in between the steep and flat DGP results.

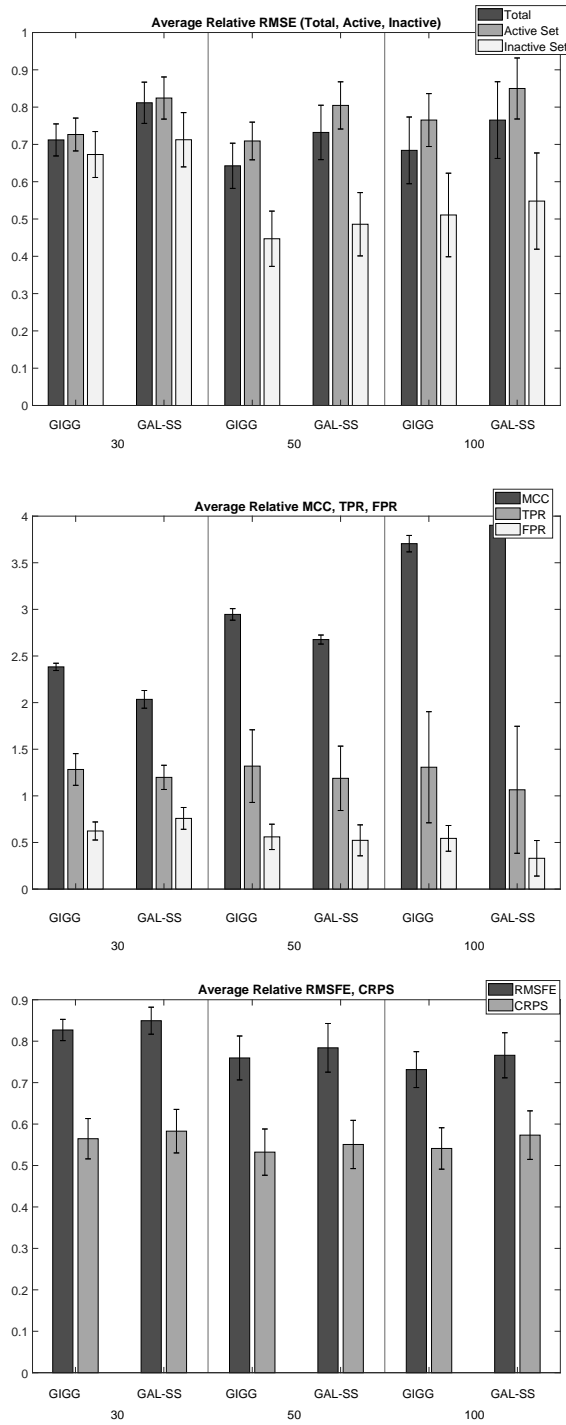


Figure 3.9: Average relative performance for a) the parameter estimation precision ($RMSE$, $RMSE(\mathcal{A})$, $RMSE(\mathcal{A}^C)$), b) variable selection (MCC, TPR, FPR) as well as nowcasting (RMSFE, CRPS). Each of these metrics are relative to those of the horseshoe prior benchmark. The horizontal axis of each panel is grouped into the two group-priors and by the covariate dimension $K = \{30, 50, 100\}$.

Table 3.3: Monte Carlo Simulations: Sparse Coefficient Vector

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

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

Results of the Horseshoe Prior

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

The effect of larger correlation comes at no surprise, as it becomes increasingly hard for shrinkage priors to distinguish between signals and noise, leading to increases in false selection of inactive groups (see the FPR columns of both tables). This causes a general decrease in the quality of variable selection (column MCC) and an increase in RMSE, where RMSE for the inactive set (column $RMSE(\mathcal{A}^C)$) increases more than for the active set (column $RMSE(\mathcal{A})$). It should be noted,

though, that the level of RMSE for the active set is always larger than that for the inactive one, independent of the DGP. This is a common finding for shrinkage priors (Zhang et al., 2022; Cross et al., 2020; Kohns and Bhattacharjee, 2022), which similar to frequentist lasso methods, retain larger non-zero bias in finite samples for active coefficients than for inactive ones.

It is therefore unsurprising to find also that the horseshoe prior struggles more with dense DGPs (first panel, table 3.4) than with sparse DGPs, showing higher RMSE for active, but also inactive sets, along with much lower MCC. Lower MCC is driven by larger FPR. And since the TPR is also higher with dense DGPs, this suggests that the horseshoe prior over-selects groups into the model.

The response to increasing the dimension of the covariate set, K , has a differing effect, most notably depending on the degree of sparsity. While table 3.3 shows either stable RMSE (flat DGP, 4th panel) or a decrease in RMSE (steep DGP, 1st panel), the dense DGPs in table 3.4 show clear increases in overall estimation error (flat and steep). Similar findings are reported Mogliani and Simoni (2021) (and references therein) for sparse DGPs and is related to the fact that for those simulations, the extra inactive covariate groups are less and less correlated with increasing K by design, for which estimation error will be lower. Thus total RMSE may even decrease in sparse settings with larger K ²⁴.

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

Finally, when comparing flat to steep DGPs, we find that RMSE is generally lower for the former, despite variable selection (see MCC) being worse. This is likely due to the flat coefficients already being closer to zero compared to steep DGPs such that falsely eliminating active groups has relatively less effect on overall RMSE. In line with this, the true positive rate is far lower for flat than for the steep DGPs. A complicating factor for the flat DGP in general is that the end-point restrictions falsely force the estimated profiles to end smoothly at zero, and hence, lower TPR compared to steep DGPs, for which the tail restrictions are a good fit, are somewhat expected²⁵.

²⁴ Similarly, variable selection suffers less in sparse than in dense DGPs as judged by the MCC.

²⁵ Using U-MIDAS in these DGPs could allow the prior to find more suitable weights for these DGPs, however at the cost of increased dimensionality.

Results of the Group Priors

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

Comparing the GIGG to the GAL-SS, figure 3.9 highlights that the GIGG almost always outperforms the GAL-SS, with the differences remaining remarkably constant across evaluation metrics and covariate sizes. The patterns that emerge are that 1) the GIGG shows relatively better performance in the active coefficient sets than in the inactive ones, 2) GIGG has overall better variable selection as gauged by MCC in $K = \{30, 50\}$, and 3) slightly better nowcasting performance. With respect to variable selection, we find that out-performance in MCC is driven by higher TPR, albeit the GIGG incurring slightly larger FPR. Hence, the GAL-SS remains more conservative on average in variable selection compared to the GIGG. These general trends, though, mask some interesting dynamics across individual DGPs.

Table 3.3 shows that the GIGG prior tends to outperform the GAL-SS in terms of parameter precision when correlation is high, such as the sparse, steep (3.3, third panel) and sparse, flat DGPs (3.3, sixth panel), as well as in dense DGPs more generally (3.4). For these sets of DGPs, the lower RMSE are driven by comparatively larger decreases in errors for the active coefficients than in the inactive sets²⁶. This is likely driven by two interrelated explanations. On the one hand, spike-and-slab priors discretise the parameter space and thus can suffer from worse mixing than continuous shrinkage priors with increases in correlation (Ishwaran et al., 2005; Malsiner-Walli and Wagner, 2018; Piironen et al., 2020), and on the other, the GIGG benefits from the extra covariate level scale which helps in adapting to within group sparsity observed for the steep DGP in particular. For the dense DGPs in which the active set is relatively more important for overall RMSE, we therefore see much improved performance. As soon as correlation is low, though, the margins between the GAL-SS and GIGG tend to decrease.

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

²⁶ For flat DGPs generally, GIGG tends to consistently outperform the GAL-SS, albeit with smaller margins than for the steep DGPs.

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

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

3.6 Conclusion

In this chapter, we have proposed a new Bayesian MIDAS framework, the T-SV-t-BMIDAS model, for nowcasting quarterly GDP growth, combined with a flexible group prior and variable selection algorithm motivated by Bayesian decision theory. In an application of the model to nowcasting UK GDP growth, we have shown that our model is able to capture sharp changes in GDP growth as seen during the Covid-19 pandemic in a relatively timely manner fashion, and works well also in more tranquil times, particularly at a later stage of the data release cycle, when it flexibly draws signals from “hard” indicators rather than survey data only.

Two important insights regarding the role of model features and prior choice emerge. First, a long-run trend or t-distributed stochastic volatility, substantially improve forecast performance relative to a simple BMIDAS. Second, the new shrinkage prior enhances nowcast performance by inducing group-wise sparsity while enabling the model to flexibly shift between signals. This feature proves particularly relevant once the Covid-19 pandemic is included into the analysis, since the heterogeneous nature of the shock calls for a nowcasting framework that can flexibly shift across signals from sparse groups of indicators, rather than relying on dense signals due to aggregate macroeconomic co-movement.

Table 3.4: Monte Carlo Simulations: Dense Coefficient Vector

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

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

A simulation exercise to investigate how well the proposed combination of the GIGG prior with sparsification generalise to a host of MIDAS settings reveals that the GIGG provides substantial improvement in parameter estimation, variable selection and nowcasting accuracy compared to the horseshoe prior and even improves on frontier group shrinkage priors in high correlated as well as dense DGPs. The covariate level shrinkage scale allows to adapt flexibly also to within-group sparsity situations. The good variable selection results, particularly for dense DGPs, which is commonly found for aggregate macroeconomic data, show that the GIGG is a competitive default prior for MIDAS modelling problems.

Chapter 4

Horseshoe Prior Bayesian Quantile Regression

4.1 Introduction

Quantile regression has been an important tool in the econometricians' toolkit when estimating heterogeneous effects across the conditional response distribution since the seminal work of Koenker and Bassett (1978). In contrast to least squares methods, it estimates quantiles of the dependent variables' conditional distribution directly, which allows for richer inference than solely focusing on the conditional mean. While highly influential in the risk-management and finance literature in calculating risk measures such as VaR (the loss a portfolios value incurs at a specified probability level), it has experienced a recent surge in the macroeconomic literature to quantify downside risks of the aggregate economy to financial conditions (Adrian et al., 2019; Figueres and Jarociński, 2020; Adams et al., 2020; Hasenzagl et al., 2020; Prasad et al., 2019; Hasenzagl et al., 2020; Carriero et al., 2020b; Korobilis, 2017; Mazzi and Mitchell, 2019).

A challenge for these purposes is that sources of risk can be numerous such that simple quantile regression is often rendered imprecise or infeasible in high dimensions. While a variety of regularization and dimension reduction techniques have been proposed for macroeconomic forecasting, (Stock and Watson, 2002, 2012b; Kim and Swanson, 2014; Bai and Ng, 2008; De Mol et al., 2008), extensions of high dimensional methods, in particular Bayesian methods, applied to quantile regression, remain relatively scant.

The Bayesian quantile regression approach, as popularized by Yu and Moyeed (2001), is based on the asymmetric Laplace likelihood (ALL), which has a special connection to the frequentist quantile regression solution, in that its maximum likelihood estimates are asymptotically equivalent (under certain regularity conditions, see Sriram et al. (2013)) to traditional quantile regression with a check-loss function (Koenker, 2005). A hurdle in the Bayesian literature has been that ALL

based methods result in improper posteriors with any but non-informative or exponential Laplace priors, where the latter results in the popular Bayesian Lasso quantile regression (Li et al., 2010; Alhamzawi and Yu, 2013; Alhamzawi et al., 2012; Chen et al., 2013). The broader Bayesian shrinkage literature has shown, however, that global-local shrinkage priors such as the horseshoe (Carvalho et al., 2010) and Dirichlet-Laplace prior (Bhattacharya et al., 2016) offer asymptotic as well as computational advantages over the former methods (Bhadra et al., 2019). These methods have not yet been considered for the Bayesian quantile regression. The aim of this chapter is to bridge this gap and extend the global local prior to quantile regression.

This chapter’s primary contribution is twofold. First, we adapt the horseshoe prior of Carvalho et al. (2010) for the Bayesian quantile regression framework (BQR) of Yu and Moyeed (2001). Second, we develop an efficient posterior sampler for the quantile specific regression coefficients based on data augmentation akin to Bhattacharya et al. (2016) which speeds up computation significantly for high dimensional quantile problems.

To showcase the performance of the horseshoe BQR (HS-BQR) we provide a large scale Monte Carlo study as well as a high dimensional VaR application to U.S. GDP (often called GaR in the literature). In the Monte Carlo study we show that the proposed estimator provides more stable and at worst, similar performance compared to a variety of Bayesian lasso quantile regression methods in terms of coefficient bias and forecast accuracy. We find that, particularly, tails of the distributions are consistently better estimated by the HS-BQR which echos findings from the Bayesian VaR literature (Chen et al., 2012). In the GaR application we show that the HS-BQR produces better calibrated forecast densities compared to the Bayesian alternatives and importantly provides the best performance for lower and upper tails which makes it a powerful tool for recession probability monitoring. The framework provided in this chapter has the additional advantage that the derived algorithms can be directly applied to other global-local priors that can be expressed as scale mixture of normals.¹

In what follows, we firstly present the methodological framework of the proposed model and sampling algorithm. Following this, we will provide evidence from Monte Carlo simulations and an empirical application of the favorable performance of the HS-BQR compared to alternative methods. We conclude with further generalizations of the algorithms provided and a discussion of our results.

¹ For an overview of global-local priors see Polson and Scott (2010).

4.2 Methodology

4.2.1 Bayesian Quantile Regression

Assuming a linear model such as

$$y_t = x_t' \beta + \epsilon_t, \quad t = 1, 2, \dots, T, \quad (4.1)$$

where $\{y_t\}_{t=1}^T$ is a scalar response variable and $\{x_t\}_{t=1}^T$ a $K \times 1$ known covariate vector, the objective function of quantile regression can be expressed as the minimised sum of weighted residuals which are zero in expectation for the given quantile $p \in (0, 1)$:

$$\hat{\beta}_p = \min_{\beta} \sum_{t=1}^n \rho_p(y_t - x_t' \beta), \quad (4.2)$$

whose solution $\hat{\beta}_p$, is a $K \times 1$ quantile specific coefficient vector. Note that the expected quantile $\hat{Q}_p(Y|X) = X \hat{\beta}_p$ is a consistent estimator of $Q_p(Y|X)$, independent of any parametric assumption about residuals $\{\epsilon\}_{t=1}^T$ (Koenker, 2005). We will maintain the assumption throughout that chapter that the design X is known. The loss function $\rho_p(\cdot)$ is often expressed as a tick loss function of the form $\rho_p(y) = [p - I(y < 0)]y$ where $I(\cdot)$ is an indicator function taking on a value of 0 or 1 depending on whether the condition is satisfied. As noted by Koenker et al. (2017), this loss function is proportional to the negative log density of the asymmetric Laplace distribution. This connection has been used to recast quantile regression as a maximum likelihood solution of model (4.1) with an Asymmetric-Laplace distribution, denoted as $\mathcal{ALD}(p, 0, \sigma)$, where σ denotes the scale of the \mathcal{ALD} . Assuming an \mathcal{ALD} error distribution, the working likelihood $f(Y|X, \beta_p, \sigma)$ becomes:

$$f(Y|\beta, \sigma) = \frac{p^T (1-p)^T}{\sigma^T} \prod_{t=1}^T \left[e^{-\rho_p(y_t - x_t' \beta_p) / \sigma} \right]. \quad (4.3)$$

4.3 remains a working likelihood since the quantile estimator is non-parametric about the error generating process (Wang, 2012). As posterior moments with conventional priors are not analytically available with an \mathcal{ALD} working likelihood, it has become standard practice in the literature to use a mixture representation, proposed by Kozumi and Kobayashi (2011), in which the \mathcal{ALD} error process can be recovered as a mixture between an exponentially distributed variable z_t with mean σ , $z_t \sim \exp(\sigma)$, and a standard normal variable, u_t , $u_t \sim N(0, 1)$:

$$\begin{aligned}
\epsilon_t &= \xi z_t + \tau \sqrt{\sigma z_t} u_t \\
\xi &= \frac{1 - 2p}{p(1 - p)} \\
\tau^2 &= \frac{2}{p(1 - p)}
\end{aligned} \tag{4.4}$$

where ξ and τ are deterministic quantile specific parameters. The conditional likelihood stacked over all observations thus becomes:

$$f(Y|X, \beta_p, Z, \sigma) \propto \det(\Sigma)^{\frac{1}{2}} \exp\left\{-\frac{1}{2}\left[(y - X\beta_p - \xi Z)' \Sigma (y - X\beta_p - \xi Z)\right]\right\}, \tag{4.5}$$

where $Y = (y_1, \dots, y_T)'$, $X = (x'_1, \dots, x'_T)'$, $Z = (z_1, \dots, z_T)'$ and $\Sigma = \text{diag}(1/(\tau^2 z_1 \sigma), \dots, 1/(\tau^2 z_T \sigma))$. Hence, the mixture representation results in a normal kernel for the likelihood which enables analytical solutions for conditional posteriors as shown below.

Throughout the chapter, we consider priors on β_p that take the following form:

$$\beta_p \sim N(\mathbf{0}_K, \Lambda_*), \tag{4.6}$$

where a prior mean of zero implies shrinkage toward sparsity and the prior variance parameters, Λ_* control the amount of shrinkage toward zero.

By applying independent priors $p(\beta_p, \sigma, Z) = p(\beta_p)p(\sigma)p(Z|\sigma)$, the conditional posterior for β_p is normal:

$$\begin{aligned}
p(\beta_p|\cdot) &\sim N(\bar{\beta}_p, \bar{\Lambda}_*^{-1}) \\
\bar{\beta}_p &= \bar{\Lambda}_*^{-1}(X'\Sigma(Y - \xi Z)) \\
\bar{\Lambda}_* &= (X'\Sigma X + \Lambda_*^{-1}).
\end{aligned} \tag{4.7}$$

The conditional posterior of the scale parameter is:

$$\begin{aligned}
p(\sigma|\cdot) &\sim \mathcal{G}^{-1}(\bar{a}, \bar{b}) \\
\bar{a} &= \underline{a} + \frac{3T}{2} \\
\bar{b} &= \underline{b} + \sum_{t=1}^T \frac{(y_t - x'_t \beta_p - \xi z_t)^2}{2z_t \tau^2} + \sum_{t=1}^T z_t,
\end{aligned} \tag{4.8}$$

where \mathcal{G}^{-1} stands for the inverse-Gamma distribution. Finally, the conditional posterior for z_t is:

$$\begin{aligned}
p(z_t|\cdot) &\sim 1/i\mathcal{G}(\bar{c}_t, \bar{d}_t) \\
\bar{c}_t &= \frac{\sqrt{\xi^2 + 2\tau^2}}{|y_t - x'_t\beta_p|} \\
\bar{d}_t &= \frac{\xi^2 + 2\tau^2}{\sigma\tau^2},
\end{aligned} \tag{4.9}$$

where $i\mathcal{G}$ stands for the inverse Gaussian density with location (\bar{c}_t) and shape (\bar{d}_t) parameter.

The conditionally conjugate posteriors (4.7,4.8,4.9) allow for efficient Gibbs sampling algorithms which for the independent prior setup have been shown to be geometrically ergodic by Khare and Hobert (2012), independent of any assumptions on X . Hence, X could include more variables than observations.

4.2.2 Shrinkage Priors

In order to efficiently estimate the posterior of a large dimensional coefficient vector in small samples, informative priors are needed. Ideally, these priors are able to separate noise variables from signals such that the noise is shrunk towards zero and signals attain their unrestricted parameter values. The global-local prior framework, initially formalized in Polson and Scott (2010), follows a particularly suitable hierarchy for such estimation problems, in which global scales dictate the overall level of sparsity, while local scales, specific to each covariate, allow to reduce or tighten the pull toward zero depending on the signal:

$$\begin{aligned}
\beta_{j,p}|\varphi_j^2, \vartheta^2 &\sim N(0, \varphi_j^2\vartheta^2), j \in (1, \dots, K) \\
\varphi_j &\sim \pi(\varphi_j), j \in (1, \dots, K) \\
\vartheta &\sim \pi(\vartheta).
\end{aligned} \tag{4.10}$$

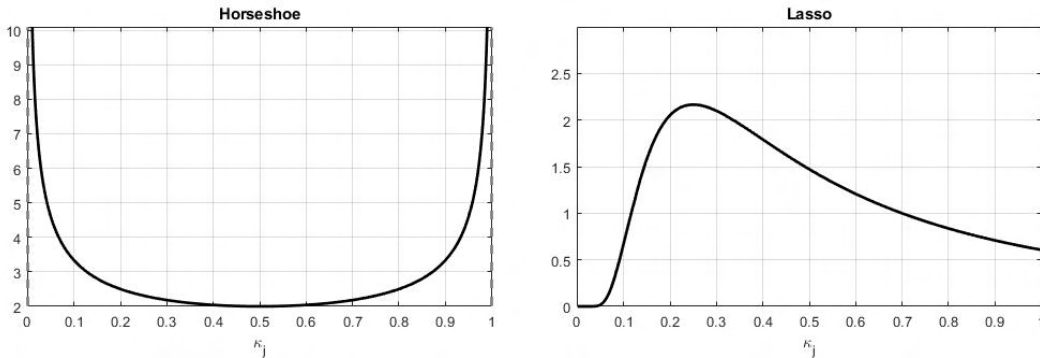


Figure 4.1: Distribution of κ_j , the shrinkage coefficient implied by the horseshoe prior and the Lasso prior.

The horseshoe prior of Carriero et al. (2020b), in particular, employs two half Cauchy distributions for φ and ϑ :

$$\begin{aligned}\varphi_j &\sim C_+(0, 1) \\ \vartheta &\sim C_+(0, 1),\end{aligned}\tag{4.11}$$

such that $\Lambda_* = \vartheta^2 \text{diag}(\varphi_1^2, \dots, \varphi_K^2)$. Notice that in our extension of the horseshoe prior to the Bayesian quantile regression, we have formulated the prior hierarchy independent of σ , while the global-local literature often uses a dependent prior. It is easily verified that the assumption of prior independence is needed so that the conditional posterior distributions are analytically available as derived above. Additionally Moran et al. (2018) have shown that in high-dimensional settings, the independence assumption aids inference of the error variance. This is due to the fact that conjugate priors act mathematically as additional observations which artificially bias the error variances downwards when $K \gg T$.

Due to that fact that half Cauchy distributions have high mass on 0 with fat tails, they are well suited for variable shrinkage tasks as they encapsulate the idea that only a few covariates are of importance. This behavior can be made explicit by slightly re-writing the posterior. Under certain conditions (see Piironen et al. (2017)), the conditional posterior for generalized linear models under any global-local prior can be reduced to the following form:

$$\bar{\beta}_j = (1 - \kappa_j)\hat{\beta}_j,\tag{4.12}$$

where $\kappa_j = 1/(1 + T\sigma^{-2}\vartheta^2 s_j^2 \varphi_j^2)$, s_j refers to the columnwise variance of X and $\hat{\beta}$ refers to the maximum likelihood estimate.² Since $\kappa_j \in [0, 1]$, the ideal behavior of the implied distribution on κ_j , which can be found by the change of variable theorem, should approximate a bi-modal with most mass on 0 or 1 depending on whether the covariate is a signal or noise. As can be visually confirmed in panel a of figure (4.1), the horseshoe prior exhibits such favorable limiting behavior, while double exponential based priors such as the lasso prior of Park and Casella (2008) do not (right panel). The lasso prior, when cast into global-local form is defined as:

$$\begin{aligned}\beta_{j,p}|\varphi_j &\sim N(0, \varphi_j), \\ \varphi_j|\phi &\sim \exp\left(\frac{\phi}{2}\right) \\ \phi &\sim \mathcal{G}(a_1, b_1),\end{aligned}\tag{4.13}$$

where $\Lambda_* = \text{diag}(\varphi_1, \dots, \varphi_K)$ and \mathcal{G} denotes the Gamma distribution. The

² For non-Gaussian models, this quantity will not exactly be equal to the maximum likelihood solution, but that of a Gaussian approximation (Piironen et al., 2017)

implied distribution on κ_j under the lasso prior has the unfortunate trait that large signals can escape shrinkage, but noise variables are not shrunk aggressively enough. Hence, this will result in too little shrinkage in large dimensional problem with many noise variables.

4.2.3 Gibbs Sampler

With the conditional posteriors at hand, we utilize a standard Gibbs sampler. The dynamics of the Markov chain $\{(\beta_m, \sigma_m, \varphi_m^2, \vartheta_m^2, z_m)\}_{m=0}^\infty$ are implicitly defined through the following steps

1. Draw $Z \sim \pi(\cdot | \beta, \sigma, \varphi^2, \vartheta^2, \theta, \tau, X, Y)$ from $1/i\mathcal{G}(\bar{c}_t, \bar{d}_t)$ for all t and call the $T \times 1$ vector z_{n+1}
2. Draw $\sigma_{n+1} \sim \pi(\cdot | \beta, \varphi^2, \vartheta^2, \theta, \tau, X, Y, z_{n+1})$ from $\mathcal{G}^{-1}(\bar{a}, \bar{b})$
3. Draw $\beta_{n+1} \sim \pi(\cdot | \sigma_{n+1}, \varphi^2, \vartheta^2, \theta, \tau, X, Y, z_{n+1})$ from $N(\bar{\beta}, \bar{V})$
4. Simulate φ_{n+1}^2 and ϑ_{n+1}^2 through slice sampling given in the appendix.
5. Iterate (1-4) until convergence is achieved

In step 4., we make use of the fact that due to the assumption of independence on the scales (φ, ϑ) , the posteriors also follow independent Cauchy distributions. Since the Cauchy has no defined moments which would enable sampling, the literature has proposed Gibbs samplers which rely either on slice sampling (Polson et al., 2014) or mixture representations (Makalic and Schmidt, 2015). Because slice sampling does not involve rejections and or extra mixing variables, we follow Polson et al. (2014) by formulating a block slice sampling algorithm for $\varphi = (\varphi_1, \dots, \varphi_K)'$ given in the appendix.

Khare and Hobert (2012) show that the Markov chain of this sampler is geometrically ergodic and also valid in $K \gg T$ settings which gives theoretical justification to apply this sampler to high dimensional settings. However, a computational bottleneck is present in very high dimensions in evaluating the $K \times K$ dimensional inverse for the conditional posterior of β . Cholesky decomposition based methods will generally be of order $O(K^3)$. Taking into consideration that in quantile settings, one is usually interested in obtaining more than one expected quantile, this can result in prohibitively long computation times. We therefore provide a more efficient sampling algorithm for β which leverages data augmentation similar to the algorithm developed by Bhattacharya et al. (2016) which is of order $O(T^2K)$ and especially beneficial in high dimensional settings.

Suppose, we want to sample from $N_K(\mu, \Theta)$, where

$$\Theta = (\Phi' \Phi + D^{-1})^{-1}, \quad \mu = \Theta \Phi' (\alpha - \xi Z). \quad (4.14)$$

Assume $D \in \mathbb{R}^{K \times K}$ is a positive definitive matrix and diagonal for simplicity, $\Phi \in \mathbb{R}^{T \times K}$, and $\alpha \in \mathbb{R}^{T \times 1}$. Then (4.7) is recovered when setting $\Phi = U^{1/2}X$, $D = \Lambda_*$ and $\alpha = U^{1/2}y$. An exact algorithm to sample from (4.7) is thus given by:

Algorithm 1. *Fast HS-BQR sampler*

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

Suppose ς is obtained through algorithm 1. Then $\varsigma \sim N(\mu, \Theta)$.

Proof. Using the Sherman-Morrison-Woodbury identity and some further algebra, $\mu = D \Phi' (\Phi D \Phi' + I_T)^{-1} (\alpha - \xi Z)$. Plugging in 2. into 3., we obtain $\varsigma = u + D \Phi' (\Phi D \Phi' + I_T)^{-1} (\alpha - v - \xi Z)$. Since by definition $v \sim N(0, \Phi D \Phi' + I_K)$, ς follows a normal distribution with mean $D \Phi' (\Phi D \Phi' + I_K)^{-1} (\alpha - \xi Z) = \mu$. As $cov(u, v) = D \Phi'$, it follows that $cov(\varsigma) = D - D \Phi' (\Phi D \Phi' + I_K)^{-1} \Phi D$ which by the Sherman-Morrison-Woodbury identity is equal to Θ . More details are provided in the appendix. The provided algorithm is not specific to the horseshoe prior and follows through for any prior of the form in (4.10). The computational advantage provided in algorithm 1 compared to Cholesky based decompositions is that we can cheaply sample from $(u, v)'$ which via data augmentation yields samples from the desired distributions. \square

4.3 Simulation Setup

In order to verify the theoretical advantages of the HS-BQR over the Laplacian based quantile regression priors laid out above, we conduct a variety of high dimensional Monte Carlo simulations that test the priors' ability to adapt to different degree's of sparsity and error distributions in the data generating processes. We consider 3 variants of the original lasso prior which have been adapted to the Bayesian Quantile regression:

1. Bayesian Lasso QR (LBQR): The lasso prior is derived by noticing that the ℓ_1 -norm penalized check loss function

$$\min_{\beta} \sum_{t=1}^T \rho_p(y_t - x'_t \beta_p) + \lambda \sum_{j=1}^K |\beta_{j,p}| \quad (4.15)$$

can be obtained as the MAP (Maximum a Posteriori) estimate of the ALL with a Laplace prior on the regression coefficients,

$\pi(\beta_p|\sigma, \lambda) = (\sigma\lambda/2)^K \exp\{-\sigma\lambda \sum_{j=1}^K |\beta_{j,p}|\}$. The posterior takes the following form:

$$\beta_p|y, X, \sigma, \lambda \propto \exp(-\sigma \sum_{t=1}^T \rho_p(y_t - x'_t \beta_p) - \sigma\lambda \sum_{j=1}^K |\beta_{j,p}|) \quad (4.16)$$

To estimate estimate (4.16), we utilize the Gibbs sampler of Li et al. (2010) with their recommended hyperpriors. Due to the shrinkage coefficient profile discussed above, we expect the LBQR to do well in sparse designs with well identified signal and noise.

2. Bayesian Elastic Net QR (BQRENET): The elastic net estimator quantile regression differs from the lasso in that it adds a ℓ_2 -norm of the regression coefficients to the minimization problem. This is the ridge component which allows to shrink coefficients in a less aggressive manner than the ℓ_1 -norm. This makes it useful when dealing with correlated or dense designs. Assuming the elastic net estimator for the quantile regression, as

$$\min_{\beta_p} \sum_{t=1}^T \rho_p(y_t - x'_t \beta_p) + \lambda_1 \sum_{j=1}^K |\beta_{j,p}| + \lambda_2 \sum_{k=1}^K \beta_{j,p}^2 \quad (4.17)$$

the prior can, similarly to above, be formulated as an exponential prior,

$\pi(\beta_{k,p}|\lambda_1, \lambda_2, \sigma) \propto \frac{\sigma\lambda_1}{2} \exp(-\sigma\lambda_1 |\beta_{j,p}| - \sigma\lambda_2 \beta_{j,p}^2)$. The posterior is then:

$$\beta_p|y, X, \sigma, \lambda \propto \exp(-\sigma \sum_{t=1}^T \rho_p(y_t - x'_t \beta_p) - \sigma\lambda_1 \sum_{j=1}^K |\beta_{j,p}| - \sigma\lambda_2 \sum_{j=1}^K \beta_{j,p}^2) \quad (4.18)$$

We use the same hyperpriors as recommended by Li et al. (2010)

3. Bayesian Adaptive Lasso QR (BALQR): The adaptive lasso as proposed by Alhamzawi et al. (2012) uses the same setup as the LBQR, but allows for the shrinkage coefficient to vary with each covariate. The prior can then be formulated as follows: $\pi(\beta_p|\sigma, \lambda_j) = (\sigma\lambda_j/2)^K \exp\{-\sigma \sum_{j=1}^K \lambda_j |\beta_{j,p}|\}$. Since this estimator allows for coefficient specific shrinkage we expect it to outperform the LBQR.

Three sample sizes are considered: $T_i \in \{200, 500\}$.³ In total 100 Monte Carlo

³ For the HS-BQR we also consider a third sample size of $T_3 = 1000$. This was done to get a better understanding whether the estimator can identify quantile varying parameters in the block design.

data sets were generated⁴ for which the last 100 observations are constructed to be the same for each T_i in order to make forecast errors comparable. The remainder of the observations are used as training samples to retrieve the mean posterior $\hat{\beta}_p$ vector to calculate bias⁵.

We consider 12 designs in total which vary along two different dimensions: the degree of sparsity and the error generating process. We test the following sparsity patterns:

- Sparse with $\beta = (1, 1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \frac{1}{5}, 0_{1 \times 2T_1})$,
- Dense with $\beta = (1, 0.85_{1 \times T_1})$,
- Block structure with $\beta = (1, 0.85_{1 \times T_1}, 0_{1 \times T_1}, 0.85_{1 \times T_1})$.

Notice that for T_1 there are always more coefficients than observations.

Consider a linear model as in (4.1). To retrieve the true quantile regression coefficients, β_p , we make use of Koenker (2005)'s alternative representation of the quantile regression:

$$y_t = x_t' \beta + (x_t' \varrho) u_t \quad (4.19)$$

where u_t is assumed to be i.i.d. having some CDF, F . The dimensionality of ϱ is $K \times 1$ and determines which covariates have non constant quantile functions. This can be seen from the solution for β_p to equation (4.19):

$$\beta_p = \beta + \varrho F^{-1}(p) \quad (4.20)$$

Hence, the true β_p profile of a quantile regression model has a random coefficient model interpretation, where the vector of coefficients can be decomposed into a fixed plus a random component. In particular, the random component depends on the inverse CDF of the error, $F^{-1}(p)$. One can therefore think of ϱ as determining which variable is correlated with the error, where by default the first entry, ϱ_0 , is set to 1. This entails that location effects will always be present.⁶

From a frequentist' perspective Equation (4.20) is our oracle estimator for β_p for a given quantile p , which, given that the ALL approximation in equation (4.5) holds, can be compared to the mean of the posterior of equation (4.7) (Kozumi and Kobayashi, 2011). With this in mind, it is trivial to calculate the true β_p 's for the error generating processes considered.

⁴ Except for $T = 500$ block case where only 20 Monte Carlo experiments were done due to the time it takes to run the estimator on such large dimensions.

⁵ Alternatively, one could also use the MAP estimate of the regression posterior as the point estimate. This might seem more natural when comparing Bayesian quantile regression methods to frequentist quantile estimators due to their equivalence as discussed in Kozumi and Kobayashi (2011). We found that since the conditional posteriors are normal, there is no practical difference between the posterior mean and MAP.

⁶ While it is possible for ϱ to take on any value, for simplicity we assume that the elements of ϱ only to take on the values $\{0, 1\}$.

Table 4.1: Summary of simulation setups

DGP	Error distributions	Quantile functions
$y_1 = X\beta + \epsilon$	$\epsilon \sim N(0, 1)$	$\beta_{0,p} = \beta_0 + F_{N(0,1)}^{-1}(p)$
$y_2 = X\beta + \epsilon$	$\epsilon \sim T(3)$	$\beta_{0,p} = \beta_0 + F_{T(3)}^{-1}(p)$
$y_3 = X\beta + (1 + X_2)\epsilon$	$\epsilon \sim N(0, 1)$	$\beta_{0,p} = \beta_0 + F_{N(0,1)}^{-1}(p)$ $\beta_{1,p} = \beta_1 + F_{N(0,1)}^{-1}(p)$
$y_4 = X\beta + \epsilon_1 + X_2\epsilon_2$	$\epsilon_1 \sim N(0, 1)$ $\epsilon_2 \sim U(0, 2)$	$\beta_{0,p} = \beta_0 + F_{N(0,1)}^{-1}(p)$ $\beta_{1,p} = \beta_1 + F_{U(0,2)}^{-1}(p)$

The second dimension along which the DGPs differ is in their error process. The proposed DGPs can be grouped into two broad cases: (1) i.i.d. errors (y_1 and y_2); and (2) heteroskedastic errors (y_3 and y_4). In y_1 we assume that the error distribution follows a standard normal distribution and in y_2 the error has student-t distributed errors with 3 degrees of freedom. For the other cases, we assume simple heteroskedasticity caused by correlation between the second covariate (whose coefficient we denote as $\beta_{1,p}$) and ϵ . Lastly, y_4 can be thought of as containing a mixture between a uniform and a standard normal error distribution. In all simulations, the design matrix is simulated using a multivariate normal distribution with mean 0 and a covariance matrix with its $(i, j)^{th}$ element defined as $0.5^{|i-j|}$.

Relating the assumed error processes to the random coefficient representation (4.20), it is clear that, under i.i.d. errors, only the constant has a non-constant quantile function caused by F^{-1} (hereinafter called location shifters). Under the heteroskedastic designs, apart from the constant, $\beta_{1,p}$ will have a non-constant quantile function as well. Hence, $\beta_{1,p}$ in y_3 is determined by $F_{N(0,1)}^{-1}$ across p , and $\beta_{1,p}$ in y_4 follows $F_{U(0,2)}^{-1}$, i.e., increases linearly with p . The simulation designs (and the true quantile functions) are summarized in table (4.1).

We evaluate the performance of the estimators in terms of bias in the coefficients and forecast error. Using the true quantile profile in β_p in (4.20), we calculate root mean coefficient bias (RMCB) and root mean squared forecast error (RMSFE) as:

1. Root Mean Coefficient Bias = $\sqrt{\frac{1}{iter} \|\hat{\beta}_p - \beta_p\|_2^2}$

2. Root Mean Squared Forecast Error = $\sqrt{\frac{1}{iter} \|X\hat{\beta}_p - X\beta_p\|_2^2}$

where $iter$ is the number of Monte Carlo experiments. For most cases $iter = 100$, except for Block T_2 , where it is set to 20.⁷

⁷ The only estimator where there is a deviation from this is the BALQR where the variance covariance matrix of the posterior coefficients was not invertible for some of the cases. This is indicative that the BALQR prior did not shrink enough

4.3.1 I.i.d. Distributed Random Error Simulation Results

The bias results for the three designs (sparse, dense, block) across a selection of quantiles are presented in table (4.2) and the results of the forecast performance are presented in table (4.3). To shed light on whether the estimators capture the variable’s quantile function appropriately, we additionally show plots for variables with non constant quantile curves for each quantile. The HS-BQR’s plots are presented in figure (4.4). The line in the plots shows the average, while the shaded region highlights the 95% coverage of β_p values across the Monte Carlo runs.

Table (4.2) shows that the HS-BQR performs competitively compared to the considered estimators in all i.i.d designs regardless of what type of sparsity structure is considered. In particular, for the sparse case the HS-BQR provides the lowest coefficient bias for both y_1 and y_2 for all quantiles. The forecast results from table (4.3) corroborate these findings with the HS-BQR providing the lowest root mean squared forecast errors among the estimators considered.

The HS-BQR’s performance is competitive for the dense and block cases as well, as can be seen in table (4.2), however falls slightly short for the central quantiles to the BQRENET in the dense and to the BALQR in the block cases for T_1 . Forecast errors in table (4.3) confirm these results. This coheres with the theoretical properties of the priors. The ridge component in the BQRENET provides better inference for dense designs, while the BALQR benefits in block structures from adaptive shrinkage without having to identify a global shrinkage parameter.

Figures (4.2) and (4.3) show the performance of the estimators at different parts of the block design for T_1 and T_2 respectively. It reveals how the HS-BQR does extremely well in the sparse regions of the data for y_1 and y_2 while not being able to replicate this performance in the dense regions of the data for T_1 . This is not to say that it performs poorly: while the HS-BQR yields higher average bias than the competing estimators, this is not statistically different from the bias of the other estimators. When more data are introduced in T_2 , the difference in bias for the dense parts become even smaller among the different priors, while the sparse parts are estimated with considerably more accuracy for the HS-BQR.

Generally, as more data are introduced, the performance across the estimators converge to similar bias and forecast results, which confirms asymptotic validity of the priors and their samplers. An exception is presented by the BALQR which seems to fare worse with more data for the block design.

Both the normally distributed y_1 and t-distributed y_2 showcase a situation where the extreme quantiles (0.1 and 0.9) have higher bias than the central quantile (0.5) for all the estimators considered. This is a common finding in quantile regressions which is on account of more extreme quantiles being “data sparse” as a few observations get large weights. While it is expected that there is a U-shape in the coefficient bias as we move across the quantiles, the slope of this shape is not uniform across

the estimators. In particular, it can be seen in table (4.2) that the HS-BQR's bias does not increase as much as the other estimators.⁸ Similarly, extreme quantiles generally tend to have higher forecast errors for all estimators, but the HS-BQR's extreme quantiles don't suffer as much as it's competition as shown in table (4.3). This property cannot be overstated, as quantile regression is often employed for extreme quantiles. The only case where the HS-BQR's extreme quantiles performance are less accurate is for the dense design of T_2 , where the BQRENET's performance does not suffer as much as the HS-BQR's when considering the extreme quantiles.

Figure (4.4) underpins the findings of the tables: the HS-BQR captures the normal inverse CDF shape for y_1 and inverse t-distribution for y_2 very well in the sparse design for all T_i , however in the dense design, it only identifies location shift's for the more extreme quantiles for T_1 . Nevertheless, this property is fixed when more data is available. The figure also highlights how the HS-BQR struggles the most with block designs: It only captures the quantile profiles correctly for T_3 . This finding underpins, that in designs with unmodeled block structures and, hence, badly identified global shrinkage, quantile effects might be shrunk away. Implementation of group-level shrinkage along with prior information about the sparsity pattern in the data might be able to alleviate this problem, which we leave for future research.

⁸ Apart from the HS-BQR in the block design of T_1 , where the estimators have lower coefficient bias and forecast error for its extreme low quantiles than its central quantiles.

Table 4.2: Root mean coefficient bias

	p	y_1									y_2									y_3									y_4								
		0.1	0.3	0.5	0.7	0.9	0.1	0.3	0.5	0.7	0.9	0.1	0.3	0.5	0.7	0.9	0.1	0.3	0.5	0.7	0.9	0.1	0.3	0.5	0.7	0.9											
T=100																																					
Sparse																																					
<i>HS-BQR</i>	0.045	0.036	0.034	0.038	0.050	0.061	0.047	0.044	0.048	0.069	0.061	0.069	0.084	0.101	0.132	0.043	0.043	0.059	0.082	0.119	0.061	0.069	0.084	0.101	0.132	0.043	0.043	0.059	0.082	0.119							
<i>LBQR</i>	0.051	0.044	0.050	0.074	0.146	0.073	0.052	0.063	0.090	0.170	4.795	2.909	7.457	3.713	2.640	4.899	2.843	7.609	3.813	2.626	4.795	2.909	7.457	3.713	2.640	4.899	2.843	7.609	3.813	2.626							
<i>BQRENET</i>	0.046	0.042	0.053	0.080	0.113	0.067	0.048	0.055	0.083	0.136	0.053	0.046	0.074	0.130	0.186	0.053	0.060	0.084	0.114	0.176	0.053	0.046	0.074	0.130	0.186	0.053	0.060	0.084	0.114	0.176							
<i>BALQR</i>	0.075	0.049	0.043	0.052	0.080	0.161	0.144	0.145	0.144	0.164	0.515	0.513	0.525	0.512	0.584	0.281	0.301	0.275	0.300	0.320	0.515	0.513	0.525	0.512	0.584	0.281	0.301	0.275	0.300	0.320							
Dense																																					
<i>HS-BQR</i>	0.711	0.710	0.709	0.716	0.722	0.721	0.722	0.721	0.727	0.738	0.767	0.763	0.764	0.771	0.773	0.764	0.766	0.774	0.780	0.786	0.767	0.763	0.764	0.771	0.773	0.764	0.766	0.774	0.780	0.786							
<i>LBQR</i>	0.780	0.731	0.728	0.773	0.816	0.782	0.741	0.721	0.773	0.849	0.811	0.759	0.753	0.807	0.871	0.764	0.726	0.742	0.778	0.838	0.811	0.759	0.753	0.807	0.871	0.764	0.726	0.742	0.778	0.838							
<i>BQRENET</i>	0.739	0.676	0.679	0.716	0.781	0.746	0.700	0.694	0.735	0.790	0.752	0.714	0.684	0.772	0.815	0.733	0.683	0.678	0.703	0.791	0.752	0.714	0.684	0.772	0.815	0.733	0.683	0.678	0.703	0.791							
<i>BALQR</i>	1.271	1.233	1.250	1.246	1.265	1.276	1.245	1.240	1.267	1.286	1.307	1.287	1.283	1.287	1.309	1.268	1.260	1.254	1.264	1.287	1.307	1.287	1.283	1.287	1.309	1.268	1.260	1.254	1.264	1.287							
Block																																					
<i>HS-BQR</i>	0.747	0.752	0.754	0.760	0.766	0.752	0.754	0.760	0.762	0.769	0.760	0.756	0.764	0.757	0.773	0.668	0.665	0.666	0.670	0.677	0.760	0.756	0.764	0.757	0.773	0.668	0.665	0.666	0.670	0.677							
<i>LBQR</i>	0.821	0.737	0.716	0.783	0.870	0.831	0.743	0.704	0.773	0.879	0.803	0.717	0.708	0.750	0.858	0.766	0.713	0.708	0.799	0.863	0.803	0.717	0.708	0.750	0.858	0.766	0.713	0.708	0.799	0.863							
<i>BQRENET</i>	0.776	0.690	0.696	0.730	0.835	0.790	0.706	0.689	0.739	0.847	0.700	0.693	0.692	0.742	0.845	0.749	0.706	0.699	0.744	0.818	0.700	0.693	0.692	0.742	0.845	0.749	0.706	0.699	0.744	0.818							
<i>BALQR</i>	0.682	0.669	0.671	0.670	0.687	0.680	0.668	0.666	0.670	0.686	0.687	0.677	0.679	0.678	0.699	0.687	0.679	0.682	0.682	0.703	0.687	0.677	0.679	0.678	0.699	0.687	0.679	0.682	0.682	0.703							
T=400																																					
Dense																																					
<i>HS-BQR</i>	0.136	0.116	0.112	0.115	0.133	0.216	0.158	0.151	0.158	0.219	0.409	0.328	0.313	0.332	0.406	0.136	0.122	0.132	0.154	0.188	0.409	0.328	0.313	0.332	0.406	0.136	0.122	0.132	0.154	0.188							
<i>LBQR</i>	0.118	0.105	0.100	0.103	0.118	0.184	0.143	0.132	0.143	0.184	0.313	0.267	0.255	0.267	0.316	0.180	0.161	0.156	0.161	0.182	0.313	0.267	0.255	0.267	0.316	0.180	0.161	0.156	0.161	0.182							
<i>BQRENET</i>	0.106	0.100	0.096	0.100	0.110	0.177	0.140	0.129	0.139	0.179	0.289	0.245	0.231	0.243	0.291	0.168	0.150	0.144	0.150	0.171	0.289	0.245	0.231	0.243	0.291	0.168	0.150	0.144	0.150	0.171							
<i>BALQR</i>	0.111	0.100	0.096	0.100	0.110	0.195	0.151	0.138	0.150	0.196	0.355	0.290	0.272	0.291	0.359	0.185	0.163	0.156	0.163	0.188	0.355	0.290	0.272	0.291	0.359	0.185	0.163	0.156	0.163	0.188							
Block																																					
<i>HS-BQR</i>	0.487	0.486	0.486	0.490	0.490	0.498	0.498	0.498	0.501	0.504	0.537	0.540	0.541	0.542	0.541	0.498	0.497	0.498	0.498	0.502	0.537	0.540	0.541	0.542	0.541	0.498	0.497	0.498	0.498	0.502							
<i>LBQR</i>	0.560	0.536	0.544	0.549	0.554	0.558	0.543	0.536	0.549	0.570	0.567	0.566	0.533	0.574	0.576	0.532	0.557	0.543	0.559	0.558	0.567	0.566	0.533	0.574	0.576	0.532	0.557	0.543	0.559	0.558							
<i>BQRENET</i>	0.513	0.504	0.507	0.518	0.514	0.527	0.526	0.506	0.519	0.534	0.534	0.558	0.537	0.527	0.547	0.505	0.497	0.497	0.509	0.526	0.534	0.558	0.537	0.527	0.547	0.505	0.497	0.497	0.509	0.526							
<i>BALQR</i>	0.837	0.847	0.830	0.846	0.846	0.845	0.850	0.828	0.852	0.832	0.878	0.866	0.857	0.863	0.863	0.857	0.843	0.841	0.832	0.844	0.878	0.866	0.857	0.863	0.863	0.857	0.843	0.841	0.832	0.844							

Table 4.3: Root mean squared forecast error

p	y_1									y_2									y_3									y_4								
	0.1	0.3	0.5	0.7	0.9	0.1	0.3	0.5	0.7	0.9	0.1	0.3	0.5	0.7	0.9	0.1	0.3	0.5	0.7	0.9	0.1	0.3	0.5	0.7	0.9											
T=100																																				
Sparse																																				
<i>HS-BQR</i>																																				
<i>LBQR</i>																																				
<i>BQRENET</i>																																				
<i>BALQR</i>																																				
Dense																																				
<i>HS-BQR</i>																																				
<i>LBQR</i>																																				
<i>BQRENET</i>																																				
<i>BALQR</i>																																				
Block																																				
<i>HS-BQR</i>																																				
<i>LBQR</i>																																				
<i>BQRENET</i>																																				
<i>BALQR</i>																																				
T=400																																				
Dense																																				
<i>HS-BQR</i>																																				
<i>LBQR</i>																																				
<i>BQRENET</i>																																				
<i>BALQR</i>																																				
Block																																				
<i>HS-BQR</i>																																				
<i>LBQR</i>																																				
<i>BQRENET</i>																																				
<i>BALQR</i>																																				

4.3.2 Heteroskedastic Error Simulation Results

As with the homoskedastic DGPs, we see that for all estimators, the error rate increases when moving away from the central quantiles and that coefficient bias as well as forecast accuracy worsens for dense and block designs compared to the sparse design. Further, the bias and forecast results in tables (4.2) and (4.3) show that the HS-BQR provides competitive performance to the alternative estimators, where it consistently outperforms the other priors for y_4 in sparse designs.⁹ Similar to the previous discussion, the HS-BQR stands out in that it provides consistently more stable inference of extreme quantiles independent of the sparsity structure, with the exception of T_2 dense.

In dense designs, as for the homoskedastic simulation results, the BQRENET aided by the ridge component in the prior, provides lower coefficient bias and forecast error, than the HS-BQR, whereas in block DGPs, the BALQR outperforms the HS-BQR for y_3 in T_1 but not in T_2 . A different picture emerges for y_4 . Here, the HS-BQR's performance in coefficient bias is only rivaled by the BQRENET for both T_1 and T_2 for both dense and block cases.

Consulting figures (4.2) and (4.3) for y_3 and y_4 shows how the HS-BQR performs particularly well in the sparse regions of the data, as was the case for y_1 and y_2 . Just like in the homoskedastic designs, the HS-BQR is not able to replicate its performance in the dense regions of the data for T_1 , but it does not do much worse than the competing estimators. Similarly, the HS-BQR's bias for sparse parts of the block DGP's are far smaller while for the dense parts, it's on par with the other estimators.

The plots in figure (4.6) provide another explanation as to why the HS-BQR's forecast performance is much better for the block case of y_4 , which is that it captures some aspects of the quantile function for $\beta_{0,p}$, even for the smallest data setting T_1 . The plots in figure (4.5) and (4.6) also highlight why the estimators have lackluster performance for y_3 and y_4 for T_1 even for the sparse designs: The estimators have difficulties identifying the quantile profiles of $\beta_{0,p}$ and $\beta_{1,p}$ simultaneously. This deficiency is amended with more data as shown by the plots for T_2 and T_3 : The HS-BQR captures the quantile profiles for both the sparse and dense DGPs, however, its performance on the block design only gets better for T_3 . This shows the scale at which the methods require data to identify the correct quantile profiles of the variables when the DGP contains mixed sparsity structures. This shows the scale at which the methods require data to identify the correct quantile profiles of the variables when the DGP contains mixed sparsity structures.

The simulations have shown that the HS-BQR provides competitive results but also that all quantile methods under consideration have difficulty simultaneously

⁹ The LBQR does surprisingly poorly in the sparse T_1 heteroskedastic cases. This is on account of the estimator completely missing the quantile profiles for both y_3 and y_4 (see respective figures in the appendix).

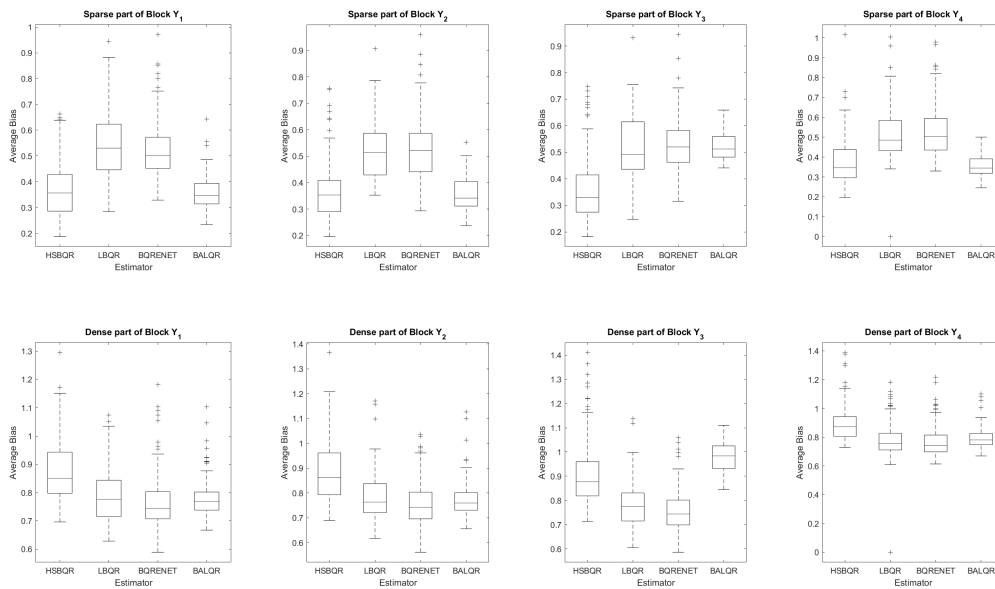


Figure 4.2: $\beta_{0,p}$ profiles for y_1 across quantiles for the different sparsity settings

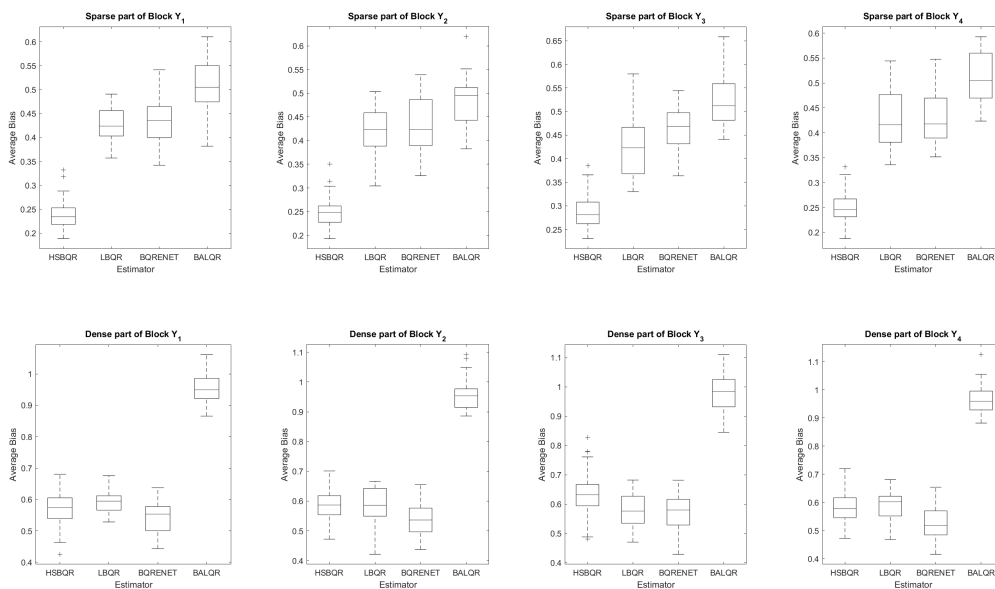


Figure 4.3: $\beta_{0,p}$ profiles for y_2 across quantiles for the different sparsity settings

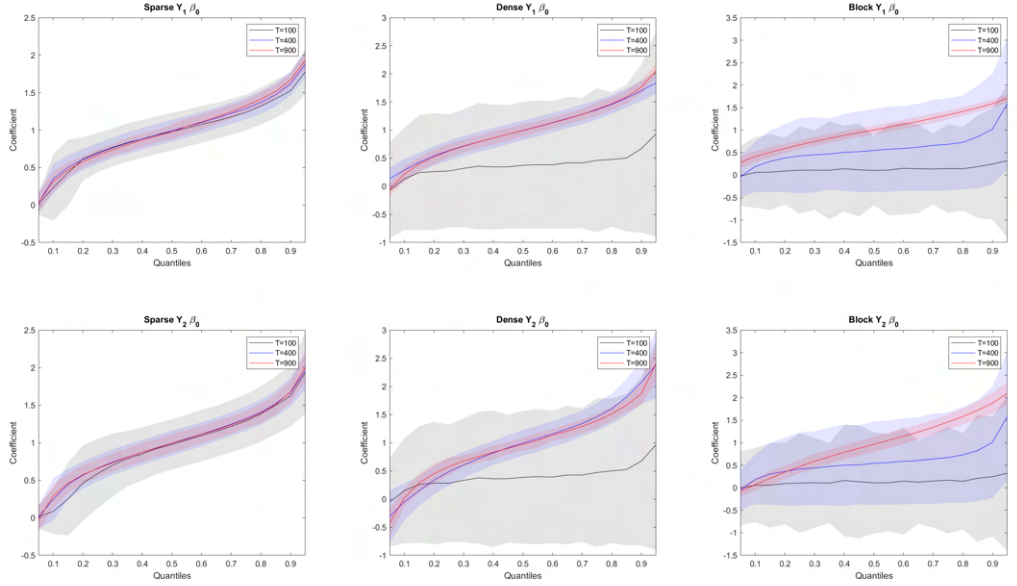


Figure 4.4: β_0 profiles for y_1 and y_2 across quantiles for the different sparsity settings

identifying the true regressors and partialling out the location ($\beta_{0,p}$) and scale ($\beta_{1,p}$) effects in high dimensional setting especially when data are not abundant.

4.4 Growth at Risk Application

We now compare the HS-BQR to the same set of competing estimators as above in estimating forecast densities of US quarterly GDP growth as well as its down- and upside risks, commonly referred to as GaR. The need for GaR was highlighted by the global financial crisis which showed how downside risks, so the lower quantiles of the density of GDP growth, evolve with the state of credit and financial market (Adrian et al., 2019; Prasad et al., 2019). Quantifying this vulnerability is of key interest of policymakers, as it is a well-known that recessions caused by financial crises are often more severe than ordinary recessions (Jordà et al., 2015).

Unlike much of the previous GaR literature which focuses on GDP growth density forecasts based on only one indicator of financial distress, we apply the HS-BQR to forecasting the entire conditional GDP density using the McCracken database, a large macro economic data set. The ability to produce well calibrated density forecasts in the face of large data contexts is important for nowcasting applications, in which the information flow is necessarily high-dimensional, or variable selection of large amount of competing uncertainty indexes. The latter purpose has been suggested by Adams et al. (2020) and Figueres and Jarociński (2020) who have argued that is not a-priori clear which index of market frictions impacts GDP growth the most.

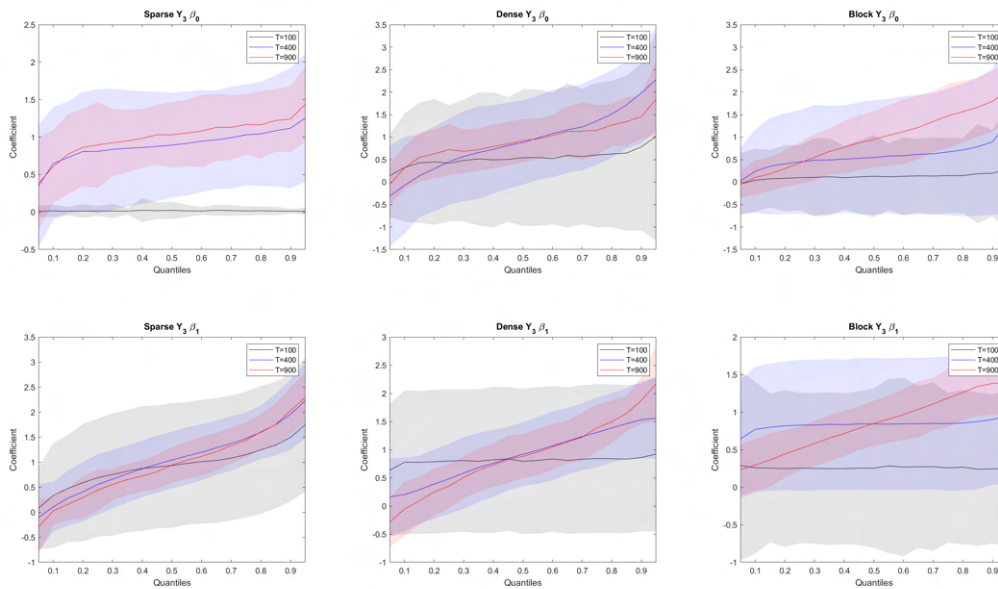


Figure 4.5: β_0 and β_1 profiles for y_3 across quantiles for the different sparsity settings

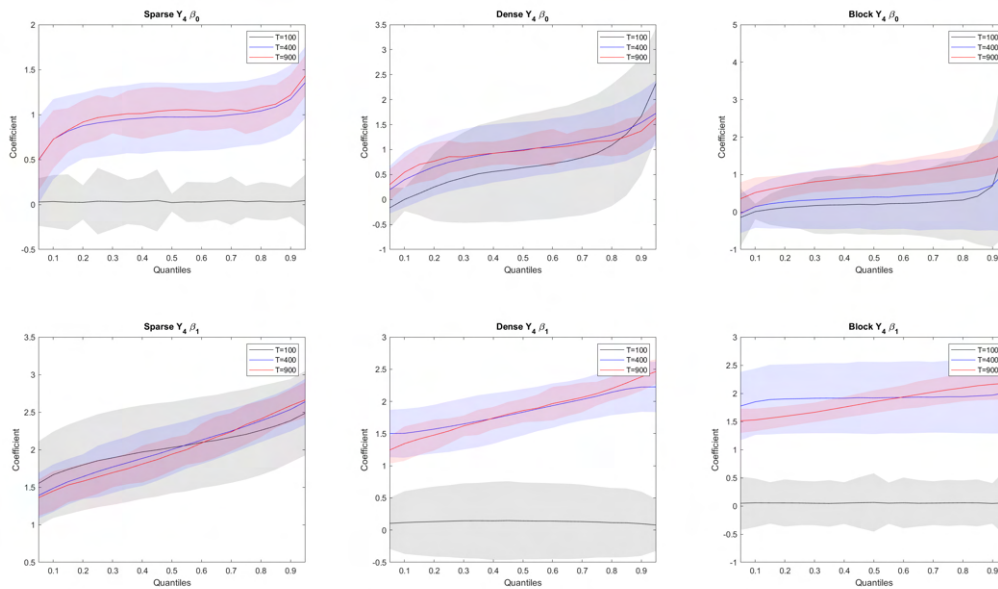


Figure 4.6: β_0 and β_1 profiles for y_4 across quantiles for the different sparsity settings

The McCracken and Ng (2020) database¹⁰ consists of 248 variables (including GDP) from 1959 Q1 at a quarterly frequency and is updated monthly. We take the quarter-on-quarter growth rate of annualized real GDP as our dependent variable and all others as independent covariates. These variables include a wide variety of macroeconomic effects which cover real, financial as well as national accounts data. Since not all variables start at 1959 Q1, for the growth at risk application, only variables that are available from 1970 Q1 were considered which gives 229 explanatory variables.

To obtain the forecasts, we use the general linear model:

$$y_{t+h} = x_t' \beta_p + \epsilon_{t+h} \quad (4.21)$$

for $t = 1, \dots, T - h$, where h refers to the forecast horizon. We consider one- to four-quarter ahead forecast horizons ($h = 1, \dots, 4$). Using the quantile setup, forecasts from each quantile are denoted as $y_{T+h|T}^p$. Note, that these h -step-ahead forecasts are equivalent to the h -step-ahead p^{th} Value-at-Risk. Forecasts are computed on a rolling basis where the initial in-sample period uses the first 50 observations of the sample, which makes for 149- h rolling forecast windows. We estimate a grid of 19 equidistant quantiles to construct the predictive density $p(\hat{y}_{T+h|T})$ via a normal kernel¹¹.

Forecast densities are evaluated along Kolmogorov-Smirnov (KS) statistics based on (unsorted) Probability Integral Transforms (PIT) and average log-scores.¹² The PIT is often used when evaluating density forecasts and provides a measure of calibration which is independent of the econometricians loss function. In particular, the PIT is the corresponding CDF of the density function evaluated at the actual observation of the out-of-sample periods, y_{t+h} :

$$g_{t+h} = \int_{-\infty}^{y_{t+h}} p(u | y_{t+h}) du = P(y_{t+h} | y_t) \quad (4.22)$$

The estimated predictive density is consistent with the true density when the CDF of g_{t+h} form a 45 degree line (Diebold et al., 1998), i.e forms the CDF of a uniform distribution. Deviation from uniformity is tested via the Kolmogorov-Smirnov test.

The unsorted PIT-based test is very close to what is referred to as hit-rate test in the Value-at-Risk literature. An additional advantage of this test is that it allows one to gauge, whether quantile-crossing is a serious problem of the estimator. Ideally, we want the forecasted quantiles to be monotonically increasing. When

¹⁰ <https://research.stlouisfed.org/econ/mccracken/fred-databases/>

¹¹ Alternatively, one could follow the popular density construction approach by Adrian et al. (2019) who fit their quantiles to a skewed t-distribution. We argue when discussing the results that this approach is less flexible than the proposed approach.

¹² There are a plethora of tests to evaluate distributions based on QQ-plot of the PIT. The choice of the KS was based solely on its simplicity to compute and any other test would suffice for evaluation.

this monotonicity is violated, our estimated density is invalid. Any violation of this monotonicity in our QQ-plot is a clear indication that the estimator yields quantiles that frequently cross.

Secondly we compare density fit via average log-scores. Log-scores provide a strictly proper scoring rule in the sense of Gneiting and Raftery (2007) and take into account location, skewness and kurtosis of the forecast distribution (Gelman et al., 2013). Since quantile crossing may lead to nonsensical density forecasts, before calculating the log-scores we sort the estimated quantiles and perform kernel smoothing to obtain $p(y_{t+h})$. Average log-scores are then calculated as follows:

$$\log S_h = \frac{1}{T-h-1} \sum_{t=1}^{T-h-1} \log p(y_{t+h}|y_t) \quad (4.23)$$

We break from the forecast density literature a bit, by not exclusively focusing on testing the whole density, but also evaluating specific quantiles' performance as well. To appraise the HS-BQR compared to the alternative estimators, the pseudo R^2 for the quantiles is computed¹³, following Koenker and Machado (1999). The pseudo R^2 of the following regression is obtained from:

$$Q_{y_{t+h}}(p|V_{t+h,p}) = \beta_0 + \beta_1 V_{t+h,p} \quad (4.24)$$

where $V_{t+h,p}$ is the fitted value of of the estimator for the p^{th} quantile. Running the regression in equation (4.24) for the p^{th} quantile gives an intuitive test for the ability of the estimated fitted value to capture the dynamics we are interested in. In particular the pseudo R^2 is calculated the following way:

$$R^2 = 1 - \frac{RASW}{TASW} \quad (4.25)$$

where $RASW$ is the residual absolute sum of weighted differences, so the residuals of equation (4.24) and $TASW$ is the total absolute sum of weighted differences, so the residuals of equation (4.24), where β_1 is constrained to 0. In essence, the pseudo R^2 shows how much information $V_{t,p}$ adds to the regression compared to a Quantile regression with only a constant.

To gain a visual understanding of how the forecast densities perform over time, figure (4.7) plots in its left panel the one-step-ahead forecast densities of the HS-BQR and the right panel shows all other competing estimators. The figure highlights that the HS-BQR provides better calibration especially in the beginning period of the forecast evaluation during which upper, lower and middle quantiles span a reasonable

¹³ Since Growth-at-Risk is meant to be a VaR of growth, utilizing tests designed to test the adequacy of VaR models is a natural extension for evaluation. Two popular tests to verify the performance of a specific quantile are the DQ test of Engle and Manganelli (2004) and the VQR test of Gaglianone et al. (2011). These tests provide a principled way of testing the null hypothesis of the selected quantile being correct. However, they do not offer a comparative measure as to how much better the proposed method provides better fit for a specific quantile.

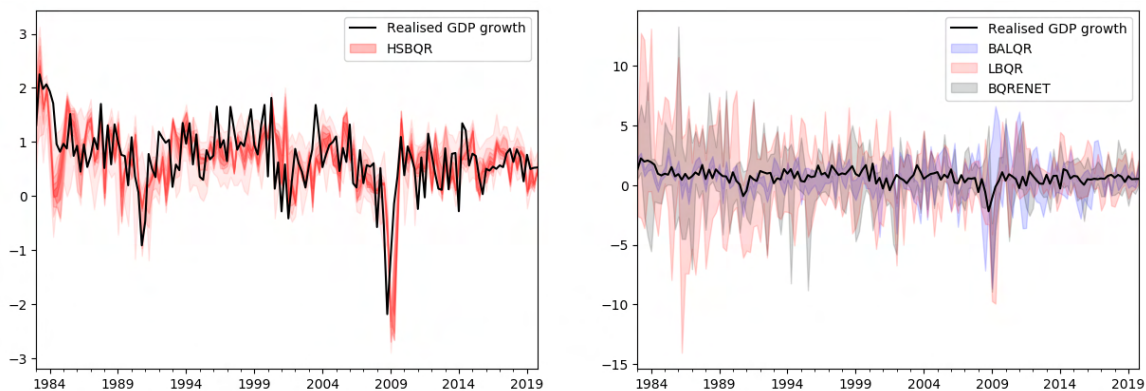


Figure 4.7: One-step-ahead forecast distributions for the L1QR, BQR, BALQR and HS-BQR. Shaded areas correspond to plots of all 19 quantiles.

range of values despite the relative scarcity of observations to number of covariates. It is clear from the right panel that precisely in the early forecast periods, the lasso based priors offer too little regularization, yielding far too extreme upper and lower quantile growth forecasts. As more data comes in, the right panel shows that the extreme variability of the lasso based estimators decreases somewhat over time, but the HS-BQR provides good fit throughout the entire evaluation period. Forecast densities for two-three- and four- quarter ahead forecast densities are presented in supplementary material and confirm these findings.

The visual inspection is corroborated by the more formal PIT based KS statistics and average forecast log-scores in table (4.4): the KS statistics show that the HS-BQR is the only estimator to provide forecasts densities whose PIT are statistically indistinguishable from a uniform CDF at the 10% significance level, and whose log-scores are highest for all but the 2-quarter ahead horizon. As expected, the test statistics as well as the QQ-plots of the PITs plotted in supplementary material, indicate that as the forecast horizon increases to 2- and 3- quarters, density calibration deteriorates somewhat for all estimators. Contrary to Carriero et al. (2020b) and Mazzi and Mitchell (2019), however, we find that density fit increases again at the 4th horizon, which suggests that the HS-BQR is useful not only for short-term density forecasts, but also for medium-term forecasts.

An additional feature of the HS-BQR forecasts is that they exhibit limited quantile crossing problem. The HS-BQR's forecasted quantiles exhibit very little quantile crossing, especially when comparing it to the alternative estimators. In fact, in the one-step-ahead case, the HS-BQR is the only estimator that yields non-crossing quantiles.

To quantify the relative performance of the estimators in capturing tail risks, we show in the third panel of table (4.4) estimates of the pseudo R^2 which are calculated as in equation (4.24) for the extreme and middle quantiles. It is apparent that not only does the HS-BQR provide better quantile fit at all shown quantiles, but that the largest margin (compared to the other estimators) is at the lowest and

Table 4.4: Performance of the different estimators for h-step-ahead quantile forecasts.

	KS	Pseudo- R^2					KS	Pseudo- R^2				
		0.05	0.25	0.5	0.75	0.95		0.05	0.25	0.5	0.75	0.95
		<i>1-step ahead</i>						<i>2-step ahead</i>				
HS-BQR	0.195	0.256	0.13	0.102	0.15	0.234	0.217	0.247	0.142	0.204	0.176	0.18
LBQR	0.382***	0.049	0.083	0.028	0.031	0.067	0.364**	0.056	0.017	0.046	0.023	0.015
BQRENET	0.396***	0.048	0.11	0.063	0.041	0.001	0.426***	0.054	0.052	0.12	0.026	0.01
BALQR	0.443***	0.11	0.025	0	0.001	0.039	0.433***	0.001	0	0.033	0.019	0.078
		<i>3-step ahead</i>						<i>4-step ahead</i>				
HS-BQR	0.247	0.158	0.117	0.147	0.101	0.188	0.193	0.296	0.13	0.132	0.111	0.071
LBQR	0.333**	0.003	0.006	0.019	0.024	0.064	0.398**	0.015	0.014	0.051	0.008	0.01
BQRENET	0.402***	0.043	0.058	0.103	0.067	0.011	0.370**	0.023	0.01	0.063	0.008	0.001
BALQR	0.491***	0.007	0.002	0.008	0.058	0.076	0.446***	0.052	0	0.017	0.035	0.088

Note: The first column, KS, reports the Kolmogorov-Smirnov test for equality to a uniform CDF with critical values of 0.374, 0.312 and 0.28 for the 1%, 5% and 10% critical values respectively (indicated by ***, ** and *). Pseudo- R^2 are shown for a set of 5 chosen quantiles.

Table 4.5: Average log scores and MSFE on the median of the different estimators for h-step-ahead quantile forecasts.

	Average Log-Scores				Median RMSFE			
	h=1	h=2	h=3	h=4	h=1	h=2	h=3	h=4
HS-BQR	3.432	3.282	3.472	3.534	0.006	0.004	0.004	0.004
LBQR	3.235	3.294	3.396	3.431	0.010***	0.012***	0.008***	0.008***
BQRENET	2.999	3.369	3.426	3.430	0.009***	0.007***	0.007***	0.008***
BALQR	1.839	2.069	2.154	2.133	0.015***	0.012***	0.010***	0.007***
SPF	3.083	3.276	3.285	3.185	0.005	0.004	0.004	0.004

Note: Density is approximated by a normal kernel of the 19 forecasted quantiles. For the MSFE stars indicate statistical difference to the HS-BQR median forecasts based on the Diebold-Mariano test (1998) at 10%, 5% and 1% significance respectively.

highest quantiles at all horizons which echos the results from the simulations. This is corroborated by the PIT graphs, which show that the HS-BQR’s tail quantiles are consistently the closest to the ideal 45-degree line.

The proposed estimator also provides competitive point forecasts which are shown in the right panel of table (4.5) for the 50th quantile. As shown by He et al. (1990), median quantile forecasts are more robust to outliers than conditional mean forecasts. Table (4.5) clearly shows that the HS-BQR offers sizable improvements in root-mean-squared-forecast-error over the competing quantile models of 25%-66%, which are all statistically significant as per the Diebold and Mariano (2002) test.

To showcase how these improvements translate to actual events of importance to policymakers, we plotted density forecasts at all horizons right before NBER marked recession or trough dates. We concentrate on the quarters before the height of each individual crisis, as the recent growth-at-risk literature highlights the usefulness of quantile methods to detect vulnerabilities to parts of the economy before these vulnerabilities materialize (Adrian et al., 2019). Representative for all other pre-crisis period shown in supplementary material, figure (4.8) shows forecast densities for Q2 2008. The actual realization is marked by a vertical grey line. Two points emerge from this graph: the HS-BQR provides the largest mass at the actual realization of growth (which translates to the highest density fit for this realization as measured by the log-score) and it provides a bi-modal distribution which yields a policy relevant characterisation of forecasted risk. The second mode hovers over negative growth outcomes, thereby giving a clear indication of risks of a recession. Compared to the HS-BQR, the competing quantile methods do provide mass on negative growth outcomes which is corroborated by (Carriero et al., 2020b; Mazzi and Mitchell, 2019) however, provide little, or close to no mass on the actual realization. In fact, consulting figure (4.7), one can see that the lasso based BQR methods throughout the entire forecast evaluation period provide mass on negative growth outcomes, in other words forecast positive probability of recessions. This is less confidence inspiring than the forecast densities of the HS-BQR which are more conservative with mass on negative growth outcomes. To argue that this is not an artifact of the kernel smoothing, we provide forecast densities for relatively ‘tranquil’ economic times, namely 2005Q1, in supplementary material. For these forecast densities, the HS-BQR combines to a unimodal, non-skewed, normal looking forecast density with high mass on the realization. This highlights an advantage of quantile smoothing compared to Adrian et al. (2019) approach of fitting the quantiles to a t-distribution. By smoothing the 19 forecasted quantiles via a kernel we impose no restrictions on the number of modes or degree of skewness of the combined density. Finally, to compare the utility of the HS-BQR approach to a widely used forecast density constructed by survey expectations, we plotted the Survey of Professional

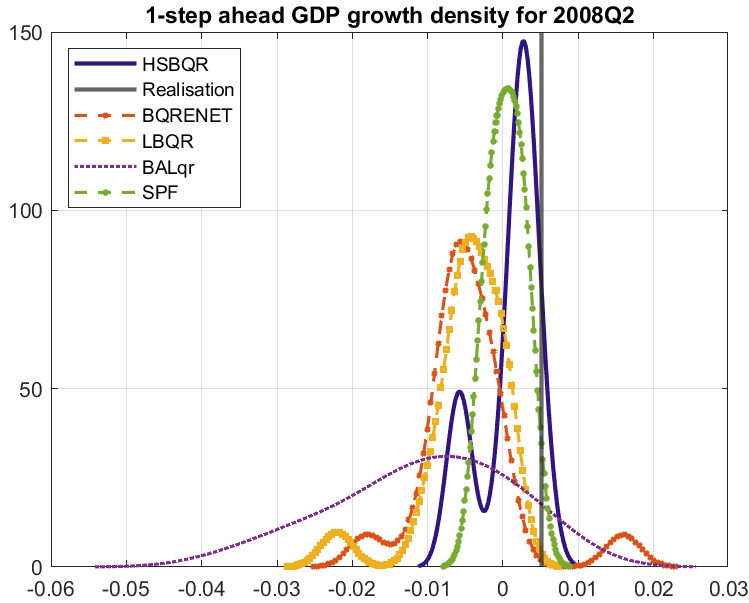


Figure 4.8: Q2 2008: Smoothed forecast densities of all competing estimators and the SPF. Densities are estimated via a Gaussian kernel of 19 equidistant forecasted quantiles. The growth realization is marked by a vertical grey line.

Forecasters distribution (SPF) into the same density graphs ¹⁴. From figure (4.8), one can see that the HS-BQR not only outperforms the SPF but provides a better indication of the looming recession indicated by larger mass on negative growth outcomes.

The ability to produce well calibrated density forecasts and, especially accurate downside risk measures in the face of large data contexts makes the HS-BQR a powerful tool for nowcasting applications or variable selection of large amount of competing uncertainty indexes.

4.5 Conclusion

In this chapter, we have extended the widely popular horseshoe prior of Carvalho et al. (2010) to the Bayesian quantile regression and provided a new algorithm to sample the shrinkage coefficients via slice sampling for the independent prior and a fast sampling algorithm that speeds up computation significantly in high dimensions.

In our simulations, we considered a variety of sparse, dense and block designs with different error distributions which revealed three points about the HS-BQR. First, the HS-BQR provides better or comparable performance in terms of both coefficient bias and forecast risk where best performance can be expected for sparse designs. Second, the aggressive shrinkage profile of the HS-BQR leads to especially good performance in tail estimation (0.1 and 0.9). Finally, an issue that all BQR

¹⁴ Smooth densities have been estimated based on a normal kernel over 19 equidistant quantiles of the survey.

methods share is simultaneously identifying the correct location and scale effects in high-dimensional setting.

Our empirical application shows that the HS-BQR provides considerable gains in calibration, density fit and even point estimates compared to double exponential based priors at all horizons, especially so at short, $h=1$. and medium term, $h=4$, horizons. Local measures of fit confirmed that HS-BQR's fitted quantiles provide the best goodness of fit. The HS-BQR proved especially useful right before NBER marked recession and trough dates, providing forecast densities foreshadowing crises. This shows that the HS-BQR is an adequate method to give credible Value-at-Risk estimates. We expect therefore that the HSBQR performs well in nowcasting settings such as (Carriero et al., 2020b; Mazzi and Mitchell, 2019) which we leave for future research.

The results show that the HS-BQR is a competitive estimator for which especially good behaviour can be expected in sparse designs with few observations. However, there are multiple fronts on which the proposed HS-BQR can be improved upon. For instance, the simulations highlighted that in dense and block designs, the HS prior tends to shrink the constant too aggressively. Hence, extensions which allow for differing shrinkage terms for subsets of the regressors might be able to alleviate this problem. Extensions to the HS-BQR should also address the problems of simultaneously estimation location and scale effects as this is needed to attain oracle properties in quantile regression.

Chapter 5

Shrink then Sparsify: Identifying Drivers of Tail Risk

5.1 Introduction

While modern day economics, and broadly social science research, is often faced with high dimensional estimation problems in which the number of potential explanatory variables is large compared to the number of observations, the extant literature for such estimation problems has focused developments to a large degree on models that characterize the conditional mean. Moving beyond the conditional mean to conditional quantiles on the other hand, has the benefit of allowing to gauge potentially heterogeneous effects of variables directly across the conditional response distribution. To this end, quantile regression has been highly influential in the risk-management and finance literature in calculating risk measures such as VaR (i.e., the loss a portfolio's value incurs at a specific probability level). Recently, the method has also experienced a surge in popularity within the macroeconomic literature to quantify risks and vulnerabilities of output growth in response to summary measures of financial health, aptly named growth-at-risk (GaR) (Adrian et al., 2019; Figueres and Jarociński, 2020; Adams et al., 2020). Quantifying this vulnerability is of key interest to policymakers, as it is well-known that recessions caused by financial crises are often more severe than ordinary recessions (Jordà et al., 2015) and thus present a welcome addition to the toolkit of policymakers.¹

Recent contributions on GaR applications such as Ferrara et al. (2020), Carriero et al. (2020b), Mitchell et al. (2022), Kohns and Szendrei (2020) (among many more), extend on the notion of estimating tail risks by considering not only financial conditions but a host of other data about the macro economy. The underlying assumption

¹ It should be mentioned that GaR should be treated distinctly from the literature on forecasting crisis periods directly through Markov-switching models (Hubrich and Tetlow, 2015; Guérin and Marcellino, 2013) or probit models (McCracken et al., 2021), as the latter predicts the probability of an event occurring, while quantile regression is typically used to model the conditional distribution of a continuous variable.

in these studies is that risks stem from many different sources which might be a-priori unknown. A common finding is that using larger data sources allows for better calibration of downside risk.

With the use of larger data sources and the limited amount of observations available for most macro time-series, Bayesian shrinkage priors such as (Li et al., 2010; Alhamzawi et al., 2012; Alhamzawi and Yu, 2013; Kohns and Szendrei, 2020; Hasenzagl et al., 2020; Korobilis, 2017) are becoming ever more pertinent to policy makers and practitioners alike. A common problem faced in these and related studies is that, while these priors often predict well, they are less apt at communicating to the policy maker which variables best characterize predictions. This stems from the a natural problem that such priors produce posteriors which are non-sparse with probability 1 (Park and Casella, 2008; Polson and Scott, 2010).

In this chapter, we propose methods to simultaneously deal with high dimensional quantile estimation and variable selection from a Bayesian decision-theoretic perspective which allows to decouple shrinkage from variable selection in the spirit of Hahn and Carvalho (2015). The goal, compared to previous implementations of variable selection in Bayesian quantile regression² is to summarize variable importance in the posterior that best describes the predictions of the un-sparsified model³. This is formulated as a decision task. We propose easy to implement adaptive sparsification procedures of frontier Bayesian quantile regression priors such as the horseshoe prior (Li et al., 2010; Kohns and Szendrei, 2020) that adapt in a data driven manner to the given quantile. Evidence from simulations as well as a high dimensional GaR application to the US shows that our procedure either preserves fit of the un-sparsified posterior or even improves both point- as well as - density fit, especially in the left tail which characterizes downside risks. Contributing further to investigating variable importance for a novel large growth-at-risk application to the US, we find evidence that there is considerable quantile specific sparsity across the conditional distribution which is distinct from the sparsity pattern found in comparable conditional mean models. In particular, we find that for the one-year ahead horizon, variables about “Interest Rates” and “Consumer Sentiment” are important for downside risk, while the right tail has higher inclusion probability for “Money and Credit” and “Non-Household Balance Sheets” information. Interestingly our method also reveals evidence for variation across time and not just across quantiles. These are in line with our understanding of the sources of recession and macro risks faced by the US economy since the 1970’s.

Communicating policy goals with stakeholders is a key component of policymaking. As such, methods that allow for clear identification of what drives downside

² Such as via spike-and-slab priors (Korobilis, 2017) who’s variable selection is driven by the marginal likelihood of discrete models

³ This in fact represents an optimal Bayesian decision task akin to the Bayesian decision theory literature going back to the seminal contribution of Lindley (1968)

(and upside) risk has the potential to be just as important for the policymaker as GaR itself. Once identified, the variables can be tracked over time, and referred to when motivating policy decisions. In this manner, the methods proposed in this chapter could help in communicating such risks.

The Literature. Before detailing the outline of the chapter, we first give a brief overview of the literature on Bayesian variable selection and shrinkage priors and how this links to valid Bayesian variable selection via a two-step approach. The canonical way for Bayesian variable selection is through Bayesian model averaging priors (Raftery et al., 1997; Clyde and George, 2004) in which the model space is discretised and posterior model probabilities allow for probabilistic statements about which is the most likely model, or to provide weights for model averaging. See Alhamzawi and Yu (2013) for an adaptation of the related g-prior approach to quantile regression. Due to combinatorial bottlenecks, much of Bayesian research in variable selection in the past decades have proposed a variety of methods (see Polson and Scott (2010) and Ishwaran et al. (2005) for excellent reviews), of which the search stochastic variable selection (SSVS) prior, dating back to George and McCulloch (1993) and George and McCulloch (1997), and adapted to the quantile regression by Korobilis (2017), has been notably influential. The SSVS prior is a discrete mixture of normals prior which stays computationally efficient by stochastically searching only among models with highest posterior probability. Despite approximating the model space, it offers good empirical performance in prediction tasks as low probability models are often ignored in the Markov chain.

Instead of approximating the discretised model space, continuous shrinkage priors on the other hand seek to include all variables at all times and instead shrink noise variables' weights toward zero so as to minimise their predictive influence. The class of global-local shrinkage priors (Polson and Scott, 2010) which include the lasso (Park and Casella, 2008), horseshoe (Carvalho et al., 2010) and Dirichlet-Laplace (DL) prior (Bhattacharya et al., 2015) increasingly supplant discrete mixture priors due to their excellent theoretical as well as empirical performance (Bhadra et al., 2019). For adaptation to the Bayesian quantile regression, see Li et al. (2010) for lasso type priors and Kohns and Szendrei (2020); Mitchell et al. (2022) for global-local priors. The continuous nature of global-local priors allows additionally for fast and efficient computational methods (Bhattacharya et al., 2016; Kohns and Szendrei, 2020).

Instead of conducting estimation and variable selection in 1 step, Hahn and Carvalho (2015) on the other hand propose a 2 step procedure in which the posterior is sparsified in an inferentially coherent way, as per Lindley (1968), via integration over squared error distance to a possible sparse and linear model which adds model size penalties. This presents an important shift away from typical SSVS type priors in that the integrated loss will often seek to minimize the distance between the pos-

terior predictions of the un-sparsified model and the simpler surrogate model. Hahn and Carvalho (2015); Ray and Bhattacharya (2018) propose the use adaptive lasso type penalisation, akin to Zou (2006) for sparsification. Piironen et al. (2020) extend this approach to Kullback-Leibler and Kowal (2021) to more general loss-functions. These approaches have been applied in numerous macroeconomic forecasting articles (see Huber et al. (2019); Kohns and Bhattacharjee (2020) among others) which show good forecasting and variable selection properties.

In section 5.2, we firstly provide a review of the BQR and shrinkage priors considered, followed by a derivation of our proposed decision theoretically motivated sparsification, and details on its implementation. In section 5.6, we conduct a large scale Monte Carlo experiment that tests the proposed methodology’s robustness to a variety of high dimensional settings. This is followed in section 5.7 by a high dimensional GaR application and lastly we conclude.

5.2 Methodology

5.2.1 Bayesian Quantile Regression

Assuming a linear model such as

$$y_t = x_t' \beta + \epsilon_t, \quad t = 1, 2, \dots, T, \quad (5.1)$$

where $\{y_t\}_{t=1}^T$ is a scalar response variable and $\{x_t\}_{t=1}^T$ a $K \times 1$ known covariate vector, the objective function of quantile regression can be expressed as the minimised sum of weighted residuals which are zero in expectation for the given quantile $p \in (0, 1)$:

$$\hat{\beta}_p = \min_{\beta} \sum_{t=1}^n \rho_p(y_t - x_t' \beta), \quad (5.2)$$

whose solution $\hat{\beta}_p$, is a $K \times 1$ quantile specific coefficient vector. Note that the expected quantile $\hat{Q}_p(Y|X) = X \hat{\beta}_p$ is a consistent estimator of $Q_p(Y|X)$, independent of any parametric assumption about residuals $\{\epsilon\}_{t=1}^T$ (Koenker, 2005). We will maintain the assumption throughout that chapter that the design X is known. The loss function $\rho_p(\cdot)$ is often expressed as a tick loss function of the form $\rho_p(y) = [p - I(y < 0)]y$ where $I(\cdot)$ is an indicator function taking on a value of 0 or 1 depending on whether the condition is satisfied. As noted by Koenker et al. (2017), this loss function is proportional to the negative log density of the asymmetric Laplace distribution. This connection has been used to recast quantile regression as a maximum likelihood solution of model (5.1) with an Asymmetric-Laplace distribution, denoted as $\mathcal{ALD}(p, 0, \sigma)$, where σ denotes the scale of the \mathcal{ALD} . Assuming an \mathcal{ALD} error distribution, the working likelihood $f(Y|X, \beta_p, \sigma)$

becomes:

$$f(Y|\beta, \sigma) = \frac{p^T(1-p)^T}{\sigma^T} \prod_{t=1}^T \left[e^{-\rho_p(y_t - x_t' \beta_p) / \sigma} \right]. \quad (5.3)$$

As posterior moments with conventional priors are not analytically available with an \mathcal{ALD} working likelihood, it has become standard practice in the literature to use a mixture representation, proposed by Kozumi and Kobayashi (2011), in which the \mathcal{ALD} error process can be recovered as a mixture between an exponentially distributed variable z_t , $z_t \sim \exp(\sigma)$, and a standard normal variable, u_t , $u_t \sim N(0, 1)$:

$$\begin{aligned} \epsilon_t &= \xi z_t + \tau \sqrt{\sigma z_t} u_t \\ \xi &= \frac{1-2p}{p(1-p)} \\ \tau^2 &= \frac{2}{p(1-p)} \end{aligned} \quad (5.4)$$

where ξ and τ are deterministic quantile specific parameters. The conditional likelihood stacked over all observations thus becomes:

$$f(Y|X, \beta_p, Z, \sigma) \propto \det(\Sigma)^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} [(y - X\beta_p - \xi Z)' \Sigma (y - X\beta_p - \xi Z)] \right\}, \quad (5.5)$$

where $Y = (y_1, \dots, y_T)'$, $X = (x_1, \dots, x_T)'$, $Z = (z_1, \dots, z_T)'$ and $\Sigma = \text{diag}(1/(\tau^2 z_1 \sigma), \dots, 1/(\tau^2 z_T \sigma))$. Hence, the mixture representation results in a normal kernel for the likelihood which enables analytical solutions for conditional posteriors as shown below.

Throughout the chapter, we consider priors on β_p that take the following form:

$$\beta_p \sim N(\mathbf{0}_K, \Lambda_*), \quad (5.6)$$

where a prior mean of zero implies shrinkage toward sparsity and the prior variance parameters, Λ_* control the amount of shrinkage toward zero.

By applying independent priors $p(\beta_p, \sigma, Z) = p(\beta_p)p(\sigma)p(Z)$, the conditional posterior for β_p is normal:

$$\begin{aligned} p(\beta_p|\cdot) &\sim N(\bar{\beta}_p, \bar{\Lambda}_*) \\ \bar{\beta}_p &= \bar{\Lambda}_*(X'\Sigma(Y - \xi Z)) \\ \bar{\Lambda}_* &= (X'\Sigma X + \Lambda_*^{-1})^{-1}. \end{aligned} \quad (5.7)$$

The conditional posterior of the scale parameter is:

$$\begin{aligned}
p(\sigma|\cdot) &\sim \mathcal{G}^{-1}(\bar{a}, \bar{b}) \\
\bar{a} &= \underline{a} + \frac{3T}{2} \\
\bar{b} &= \underline{b} + \sum_{t=1}^T \frac{(y_t - x_t' \beta_p - \xi z_t)^2}{2z_t \tau^2} + \sum_{t=1}^T z_t,
\end{aligned} \tag{5.8}$$

where \mathcal{G}^{-1} stands for the inverse-Gamma distribution. Finally, the conditional posterior for z_t is:

$$\begin{aligned}
p(z_t|\cdot) &\sim 1/i\mathcal{G}(\bar{c}_t, \bar{d}_t) \\
\bar{c}_t &= \frac{\sqrt{\xi^2 + 2\tau^2}}{|y_t - x_t' \beta_p|} \\
\bar{d}_t &= \frac{\xi^2 + 2\tau^2}{\sigma \tau^2},
\end{aligned} \tag{5.9}$$

where $i\mathcal{G}$ stands for the inverse Gaussian density with location (\bar{c}_t) and rate (\bar{d}_t) parameter.

The conditionally conjugate posteriors (5.7,5.8,5.9) allow for efficient Gibbs sampling algorithms which for the independent prior setup have been shown to be geometrically ergodic by Khare and Hobert (2012), independent of any assumptions on X . Hence, X could include more variables than observations, dependent or deterministic variables.

Since taking the inverse of the posterior covariance of the regression coefficients $\bar{\Lambda}_*$ can be computationally demanding in high dimensions, we make use of the fast BQR sampler proposed by Kohns and Szendrei (2020) that reduces the computational complexity involved in obtaining a draw from $\bar{\beta}_p$ from $\mathcal{O}(K^3)$ to $\mathcal{O}(T^2K)$. This algorithm is particularly suitable for macroeconomic data sets in which K is typically much larger than T . Details of the exact sampling steps for all priors considered are given in the appendix.

5.3 Shrinkage Priors

Shrinkage priors, both for conditional mean and quantile models, can be understood as a Bayesian generalisation to frequentist penalised regression, where the penalisation is a function of the number of active coefficients:

$$\beta = \operatorname{argmin} \sum_t h(y_t, x_t, \tilde{\beta}) + \delta Q(\tilde{\beta}). \tag{5.10}$$

h and Q are two positive functions, and δ controls the amount of penalisation. Choosing $h(\bullet)$ to be the negative log-likelihood of the data and $Q(\tilde{\beta})$ to be the ℓ_1 -norm, $\|\tilde{\beta}\|_1$, (5.10) yields sparse optimal solutions for β which are maximum likelihood equivalents to the popular lasso estimator of Tibshirani (1996) when the likelihood is Gaussian. Likewise, using ℓ_1 -norm penalisation and setting $h(\bullet)$ to the negative log of the \mathcal{ALD} likelihood, (5.10) recovers the maximum likelihood equivalent of the quantile lasso of Chernozhukov et al. (2010). From a Bayesian perspective, $Q(\beta)$ instead can be understood as the negative log prior distribution imposed on β , which renders (5.10) the negative log-posterior. The Bayesian paradigm has the added advantage of being able to set a prior also for the amount of penalisation $p(\delta)$ which gives a probabilistic way to conduct shrinkage. Different prior forms of $p(\delta)$ will result in different shrinkage properties.

Lasso Prior. Firstly introduced by Park and Casella (2008), the lasso prior generalises the frequentist lasso by placing a mixture exponential prior on δ . Adapted to the BQR by Li et al. (2010), the prior takes the following form:

$$\begin{aligned}\beta_j|\varphi_j &\sim N(0, \varphi_j), \\ \varphi_j|\phi &\sim \exp\left(\frac{\phi}{2}\right) \\ \phi &\sim \mathcal{G}(a_1, b_1)\end{aligned}\tag{5.11}$$

where $\Lambda_* = \text{diag}(\varphi_1, \dots, \varphi_K)$ and \mathcal{G} denotes the Gamma distribution. The conditional posteriors for the hyperparameters are standard:

$$\begin{aligned}p(\varphi_j^{-1}|\bullet) &\sim i\mathcal{G}\left(\sqrt{\frac{\phi}{\beta_{j,p}^2}}, \phi\right) \\ p(\phi|\bullet) &\sim \mathcal{G}\left(K + a_1, \frac{1}{2} \sum_{j=1}^K \varphi_j + b_1\right).\end{aligned}\tag{5.12}$$

If instead ϕ is allowed to vary with the same distribution as in (5.11) for each j , this prior will become the adaptive lasso prior of Alhamzawi et al. (2012)⁴.

Horseshoe Prior. Similar to the standard result as to why the ℓ_1 -norm penalty in a frequentist approach to solving (5.10) tends to over shrink signals in lasso regression (Zou, 2006), the Bayesian lasso over shrinks large coefficients as well⁵.

⁴ Since in simulations and the application we found there to be no substantial difference in terms of performance between the Bayesian lasso and adaptive lasso BQR, we report only results for the Bayesian lasso. See Kohns and Szendrei (2020) for a thorough investigation of the differences in performance between these priors.

⁵ Carvalho et al. (2010) and Piironen et al. (2017) formalise this argument through implied

The general class of global-local priors seek to solve this problem by adding another shrinkage scale hierarchy, which, with suitably chosen priors, approximate the ideal behaviour of shrinking noise variables aggressively toward zero, while leaving signal variables untouched. Global-local priors take the following form:

$$\begin{aligned}\beta_j|\varphi_j^2, \vartheta^2 &\sim N(0, \varphi_j^2\vartheta^2), j \in (1, \dots, K) \\ \varphi_j^2 &\sim \pi(\varphi_j^2), j \in (1, \dots, K) \\ \vartheta^2 &\sim \pi(\vartheta^2).\end{aligned}\tag{5.13}$$

The horseshoe prior of Carvalho et al. (2010) employs two independently distributed half Cauchy distributions on the positive support for both the global and local scales:

$$\varphi_j \sim C_+(0, 1)\tag{5.14}$$

$$\vartheta \sim C_+(0, 1),\tag{5.15}$$

where $\Lambda^* = \vartheta^2 \text{diag}(\varphi_1^2, \dots, \varphi_K)$. Due to the assumption of independence on the scales (φ, ϑ) , it is straightforward to show that their posteriors follow independent Cauchy distributions. To sample from their posteriors, we make use of the slice sampler used for the Horseshoe-BQR in Kohns and Szendrei (2020).

SSVS Prior. While the lasso and horseshoe prior are continuous shrinkage priors, the SSVS prior, dating back to Mitchell and Beauchamp (1988) and George and McCulloch (1993, 1997), discretises the model space akin to Bayesian model averaging priors (Raftery et al., 1997; Clyde and George, 2004). Since evaluating each model to compare marginal likelihoods becomes quickly infeasible in high dimensions, the SSVS saves computation time by exploring models through its Markov chain with highest posterior probability. This is achieved by modeling β by a mixture prior, where coefficients are sorted into two groups, the ‘‘spike’’ and the ‘‘slab’’. When sorted into the first group, the value of the coefficient is shrunk heavily toward zero, while in the second it is modelled through a disperse normal prior. We follow George et al. (2008)’s implementation, which has been adapted to the BQR by Korobilis (2017):

$$\begin{aligned}\beta_{j,p}|\gamma_j, \varphi_j &\sim (1 - \gamma_j)N(0, c\varphi_j^2) + \gamma_jN(0, \varphi_j^2)\forall j \in \{1, \dots, K\} \\ \varphi_j^2 &\sim \mathcal{G}(a_2, b_2) \\ \gamma_j|\pi_0 &\sim \text{Bern}(\pi_0) \\ \pi_0 &\sim \mathcal{B}(a_3, b_3),\end{aligned}\tag{5.16}$$

where $\text{Bern}(\bullet)$ stands for the Bernoulli distribution, $\mathcal{B}(\bullet)$ for the Beta distribution and $c = 10^{-5}$, which effectively shrinks the spike group of coefficients to 0. π_0

shrinkage distributions for the normal means and normal regression model respectively.

controls the probability of inclusion into the slab group. Λ^* for the SSVS priors becomes $diag(\varphi_1^2, \dots, \varphi_K^2)$ for all j if $\gamma_j = 1$ and $diag(c\varphi_1^2, \dots, c\varphi_K^2)$ if $\gamma_j = 0$. Under prior (5.16), the conditional posteriors are:

$$\begin{aligned}\varphi_j^2 | \bullet &\sim \mathcal{G}(a_2 + \frac{1}{2}, \frac{\beta_{j,p}^2}{2} + b_2) \\ \gamma_j | \bullet &\sim \text{Bern}\left(\frac{\pi_0 N(\beta_{j,p}, \varphi_j^2)}{\pi_0 N(\beta_{j,p}, \varphi_j^2) + (1 - \pi_0) N(\beta_{j,p}, c\varphi_j^2)}\right) \\ \pi_0 | \bullet &\sim \mathcal{B}(1 + a_3, k - 1 + b_3),\end{aligned}\tag{5.17}$$

where k denotes the size of the slab group. Due to the very strong shrinkage implied by the scalar c , we treat $\mathbb{1}'\gamma_j^s | \bullet$ as a posterior estimate of the model size on an iteration basis. Further, throughout all simulation and the empirical application, we set $a_3 = b_3 = 1$, which embeds the assumption that a-priori all model sizes are equally likely, and thus allows for dense as well as sparse models as recommended by Giannone et al. (2021).

5.4 Decoupling Shrinkage and Sparsity

While continuous shrinkage priors such as global-local priors in (5.13) yield good forecasting performance (Cross et al., 2020; Huber et al., 2019; Kohns and Bhattacharjee, 2022), interpretation of forecasts based on the posterior are impeded by the fact that the marginal posterior $p(\beta_p | Y)$ is continuous on \mathbb{R}^K . To aid interpretability, the assumption of sparsity is often employed, which forces small coefficients to zero when they have small enough effects on the target, Y . The Bayesian approach to enforcing sparsity can be seen as an optimal action that minimises an expected loss function that embeds the preference of sparsity as:

$$\mathcal{L}(\tilde{Y}, \psi) = \zeta \|\psi\|_0 + T^{-1} \|X\psi - \tilde{Y}\|_2^2,\tag{5.18}$$

which was proposed by Hahn and Carvalho (2015), and $\|\bullet\|_0$ refers to the ℓ_0 -norm. This loss function also offers the interpretation of a policy maker's utility function for the best possible linear surrogate model whose predictions resemble those from the un-sparsified model (measured in ℓ_2 -distance) with a preference for interpretability, which is encoded as exact sparsity inducing penalization. \tilde{Y} refers to a realisation of the predictive distribution with density:

$$p(\tilde{Y} | Y) = \int p(Y | \tilde{X}, \Upsilon) p(\Upsilon | Y, \tilde{X}) d\Upsilon.\tag{5.19}$$

In (5.19), we have collected all unknown parameters into Υ for the sake of brevity. For simplicity, we focus here on in-sample predictions which renders \tilde{X} known and

equal to X .⁶ Note that in the following discussion, we omit the condition on \tilde{X} for readability. Since \tilde{Y} is a latent quantity, the expectation of (5.18) requires integration both with the respect to \tilde{Y} and Υ :

$$\mathcal{L}(\psi) = \int \int \zeta \|\psi\|_0 + T^{-1} \|X\psi - \tilde{Y}\|_2^2 p(\tilde{Y}|\Upsilon) d\tilde{Y} p(\Upsilon|\tilde{Y}) d\Upsilon. \quad (5.20)$$

Theorem 1: Assume the observation likelihood is the ALD, given by (5.3), and that the posterior for Υ , $p(\Upsilon|Y)$, given by (5.7), (5.8), and (5.9), has been obtained through posterior sampling, then the expected loss (5.20) is given by:

$$\mathcal{L}(\psi) \propto \zeta \|\psi\|_0 + T^{-1} \|X\psi - X\bar{\beta}_p\|_2^2 + \text{tr}(X'X\bar{\Sigma}_{\beta_p}) - T^{-1}\psi'X'\xi\bar{Z}_p, \quad (5.21)$$

where $\bar{\beta}$ and $\bar{\Sigma}_{\beta_p}$ refer to the posterior mean and covariance of $p(\beta_p|Y)$, and \bar{Z}_p is defined as $\bar{Z}_p = \left(\frac{|y_1 - x_1'\beta_p|}{\sqrt{\xi^2 + 2\tau^2}} + \frac{\sigma\tau^2}{\xi^2 + 2\tau^2}, \dots, \frac{|y_T - x_T'\beta_p|}{\sqrt{\xi^2 + 2\tau^2}} + \frac{\sigma\tau^2}{\xi^2 + 2\tau^2} \right)'$. The proof is provided in the appendix.

Remark on Theorem 1: the integrated loss function (5.21) is remarkably similar to the normal likelihood case derived in Hahn and Carvalho (2015), with the difference being a term involving the posterior means of Z , \bar{Z}_p , which appear due to the mixture representation of the ALD. Since we are interested in sparsifying the vector $\beta_p|_{\bullet}$, so as to minimise the Euclidean distance to the expected quantile $\hat{Q}_p = X\beta_p|_{\bullet}$, instead of the expected location of the ALD directly, we proceed from this analysis by neglecting the term involving \bar{Z}_p . While this may seem a strong simplification, notice that 1) the contributions to this term become very small with increasing T as the entries of \bar{Z}_p are multiplied one covariate at a time, which will become clear from the coordinate descent algorithm employed below; and 2) for $p = 0.5$, the term vanishes completely due to $\xi = 0$. Thus central quantiles are virtually unaffected. Lastly, ignoring the terms involving \bar{Z}_p can also be understood from the perspective that the ALD in the BQR is often only treated as a working likelihood rather than the true data generating process of Y in order to retrieve posterior estimates of the quantile regression coefficients (Yang et al., 2016). To convince on the negligible effect of dropping the last term in (5.21), we provide simulation evidence below.

Proceeding by dropping all constant terms, the objective function becomes:

$$\psi_p^* = \underset{\psi}{\text{argmin}} \zeta \|\psi\|_0 + T^{-1} \|X\psi - X\bar{\beta}_p\|_2^2. \quad (5.22)$$

At this point, it is interesting to note the difference in the objective function to frequentist penalised quantile regression such as the quantile lasso of Chernozhukov et al. (2010). While Chernozhukov et al. (2010) compute the expected quantiles

⁶ Note that in general, this need not be the case and X might contain an entirely different subset to \tilde{X} . See Hahn and Carvalho (2015) and Piironen et al. (2020) for more discussion on that.

from observations Y , (5.22) finds a sparsified vector ψ_p^* by minimising the squared differences to the expected quantile directly. Hence, due to the availability of the posterior for $\beta_p|\bullet$, $\bar{\beta}_p$, the quantile can be treated as observed. To verify that indeed squared error loss applied in (5.22) results in recovering the desired quantile coefficients, we provide Lemma 1 below:

Lemma 1: Let $\beta_p^* \in \mathbb{R}^K$ be the true quantile regression parameter vector, and X be of full column rank. Then, the following minimisation problem recovers the true quantile coefficients:

$$\hat{\beta}_p = \underset{\tilde{\beta}}{\operatorname{argmin}} \|X\beta_p^* - X\tilde{\beta}\|_2^2/T. \quad (5.23)$$

Proof: Expanding (5.23) and solving the first order conditions yields:

$$\hat{\beta}_p = (X'X)^{-1}X'X\beta_p^* = \beta_p^*.$$

5.5 Signal Adaptive Variable Selection for the BQR

Following Hahn and Carvalho (2015) and Ray and Bhattacharya (2018), we make three modifications to the objective function (5.22). Firstly, we make use of the ℓ_1 -norm instead of ℓ_0 -norm penalisation to obtain a convex objective function whose solution is computable with standard techniques such as the coordinate descent algorithm of Friedman et al. (2010). Secondly, it is well known that the ℓ_1 -norm penalisation as an approximation to the ℓ_0 norm overshrinks signals which might hurt predictive performance. To tackle this we make use of adaptive penalisation akin to Zou (2006):

$$\psi_p^* = \underset{\psi}{\operatorname{argmin}} \frac{1}{2} \|X\bar{\beta}_p - X\psi\|_2^2 + \sum_{j=1}^K \zeta_j |\psi|_j. \quad (5.24)$$

Lastly, Ray and Bhattacharya (2018) have observed that starting the coordinate descent algorithm with the posterior means of $\psi|\bullet$ results in convergence after the first iteration by setting $\zeta_j = \frac{1}{|\bar{\beta}_{p,j}|^\kappa}$ and fixing $\kappa = 2$. Hence, the solution to (5.24) with one iteration of the coordinate descent algorithm simplifies to:

$$\hat{\psi}_{p,j}^* = \operatorname{sign}(\bar{\beta}_{p,j}) \|X_j\|^{-2} (|\bar{\beta}_{p,j}| \cdot \|X_j\|^2 - \zeta_j)_+, \quad (5.25)$$

where $\operatorname{sign}(y) = 1$ if $y \geq 0$ and -1 otherwise, and let $y_+ = \max\{y, 0\}$. See the appendix for a full derivation of the coordinate descent algorithm.

While the original SAVS procedure proposed by Ray and Bhattacharya (2018) provides good sparsification properties by fixing $\kappa_j = 2$ across all coefficients for conditional mean models, a uniform value of κ across quantiles can be inappropriate if the degree of sparsity varies across the conditional distribution. Quantile specific

sparsity can arise when the coefficients β_p have a quantile profile which crosses the zero line for any $p \in (0, 1)$. Hence, to accommodate this, we make two contributions to the SAVS algorithm which we call the BQR_{SAVS} algorithm. Firstly, we treat κ_j as a parameter to be estimated from the data. And secondly, since standard score functions such as the log-score employed in projective variable selection (Piironen et al., 2020) or other functionals (Kowal, 2021) would entail variable selection consistent with conditional mean models, we use quantile specific score functions. In particular, we make use of the quantile BIC (qBIC) of Lee et al. (2014) and choose $\hat{\kappa}_p$ such that it minimises the qBIC criterion:

$$\begin{aligned} \hat{\kappa}_p &= \underset{\kappa}{\operatorname{argmin}} \quad qBIC(\kappa) \\ &= \underset{\kappa}{\operatorname{argmin}} \quad \log\left(\sum_{t=1}^T \rho_p(y_t - x'_t \psi_p^*(\kappa))\right) + |\hat{\mathcal{S}}_\kappa| \frac{\log(T)}{2T} C_K, \end{aligned} \quad (5.26)$$

where $|\hat{\mathcal{S}}_\kappa|$ is the cardinality of sparsified $\psi_p^*(\kappa)$ conditional on parameter $\hat{\kappa}_p$, and C_K , is a positive constant which diverges to infinity with K . We follow Lee et al. (2014) by setting $C_K = \log(p)$. Lee et al. (2014) show that estimating the level of penalisation in ℓ_1 penalised quantile regression through the qBIC, results in model selection consistency under expanding K and T .⁷ While model selection consistency is hard to prove in the Bayesian paradigm and is an active field of investigation, it stands to reason that we get close to model selection consistency in the frequentist sense when $\bar{\beta}_p$ is consistently estimated. We leave this proof for future research.

Alternatively to finding $\hat{\kappa}_p$ through minimisation of (5.26), one can embed a decision rule that incorporates uncertainty in $\hat{\kappa}_p$, such as choosing the smallest $\hat{\kappa}_p$ (and therefore the smallest model size), that is no worse as measured by the qBIC, to the un-sparsified posterior with probability ς . Obtaining a standard error estimate for the difference between the qBICs of the sparsified and un-sparsified posteriors are readily available from the the output of the Markov chain. We found, however, throughout simulations as well as the application that either method gave similar results, such that we use the minimum qBIC procedure as in (5.26). We will denote the BQR_{SAVS} using $\hat{\kappa}_p$ that minimises the qBIC as BQR_{BIC} .

The importance of selecting $\hat{\kappa}_p$ in a data informed manner, even for simple data generating processes without quantile varying sparsity, can be gauged from the convergence path of the coordinate descent algorithms, shown in figure (5.1).

Taking average solution paths over 50 simulations based on a simple high dimensional block and sparse DGP, which are further elaborated in section 5.6, figure (5.1) shows clearly that convergence is reached for the qBIC penalised SAVS objective

⁷ Model selection consistency is also dependent on specifying an upper bound, s , for \mathcal{S} when $K \gg T$. Not specifying s still leads to model selection consistency, however, with mild extra assumptions on the rate of K .

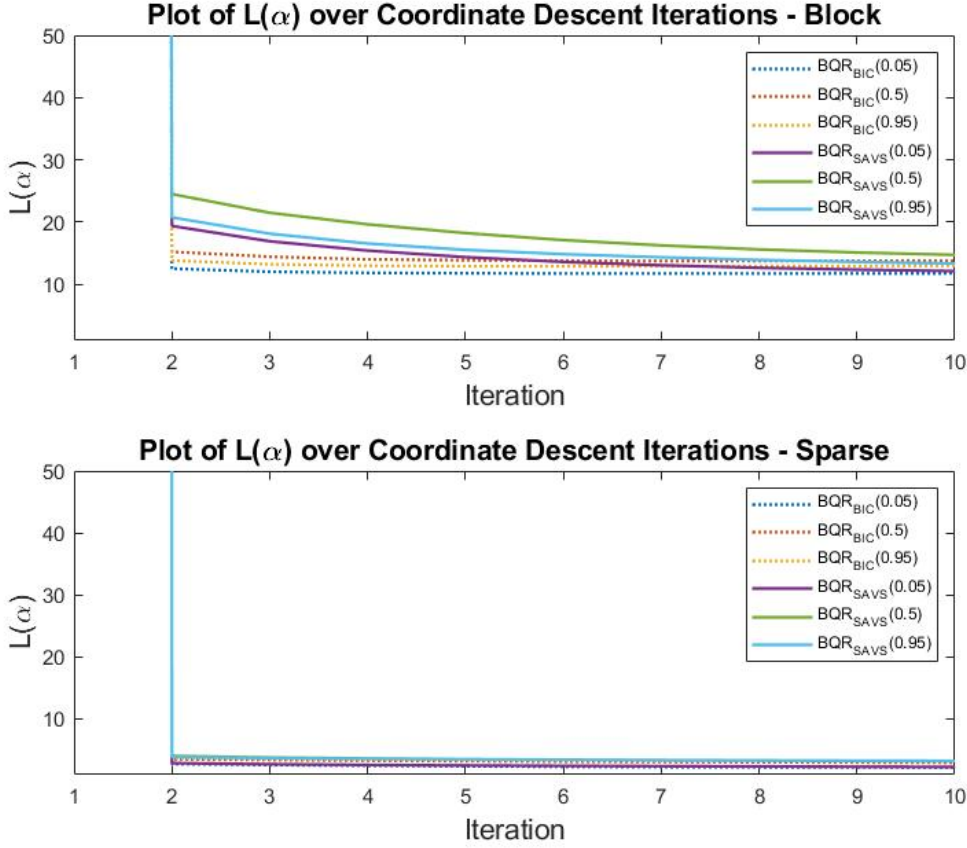


Figure 5.1: Coordinate Descent paths for objective function (5.24) for a block-sparsity and sparse DGP taken from the Monte Carlo simulations below. The convergence paths are an average over 50 simulations with 100 in-sample observations with 100 covariates, not including the constant.

function after only one iteration, while the plain SAVS takes usually more than 10 iterations on DGPs which are non-sparse. The level of the objective function is additionally much lower for qBIC compared to SAVS. Hence, stopping the coordinate descent algorithm with the plain SAVS bears the risk of inefficient penalisation.

Finally, since with correlated designs, the mean of the posterior distribution of β , can perform badly in terms of forecasting due to multimodality, we implement sparsification (5.25) as well as penalisation parameter choice (5.26) on an iteration basis. This bears the added advantage of being able to quantify variable selection uncertainty through the frequency of how often a coefficient is selected into model \mathcal{M}^s , where \mathcal{M}^s denotes a quantile regression model indexed by a vector of binary indicators of dimension K per iteration s . The percentage of times coefficient j appears in \mathcal{M}^s , can thus be interpreted akin to posterior variable inclusion probabilities such as $\mathbb{E}_\gamma(\gamma_j|\bullet)$. Alongside posterior credible intervals for ψ_p , this gives an interpretable, Bayesian decision theoretically motivated way to conduct variable selection for any continuous prior for the BQR framework.

5.6 Monte Carlo Experiment

5.6.1 Monte Carlo Setup

Simulated experiments are conducted in order to investigate the potential benefits of sparsification for the BQR in terms of its variable selection properties, coefficient bias and robustness to different data generating processes (DGPs). Another purpose of this simulation experiment is to compare variable selection of the proposed methods to the standard way of conducting Bayesian variable selection, namely via the SSVS prior. In particular, we are interested in the BQR_{SAVS} ' and BQR_{BIC} 's ability to adapt to different degrees of sparsity and error distributions. We generate data from the following model:

$$y_t = \beta_0 + x'_{t,c}\beta_c + x'_{t,q}\beta_q + (x'_t\varrho)u_t, \quad t = 1, \dots, T, \quad (5.27)$$

where $\{u_t\}_{t=1}^T$ are assumed to follow some otherwise unspecified cumulative density function (CDF), F , and the subscripts c and q refer to covariate groups whose regression coefficients don't vary and vary across quantiles respectively. From (5.27), one can see that the quantile profile of the quantile varying coefficients is in turn dictated by the correlation between the covariate and error process u_t which is enforced through ϱ being a K -dimensional binary vector which is non-zero if $x_{t,i} \in x_{t,q}$ (as well as for the constant coefficient). Assuming a tick-loss function and solving for quantile regression coefficients $\beta(p)$ for any p , Koenker (2005) has shown that quantile regression coefficients have a random coefficient interpretation, where the quantile profile is proportional to the quantiles of the error process, with its centre on β_c and β_q respectively:

$$\begin{aligned} \beta_0(p) &= \beta_0 + \varrho_0 F^{-1}(p) \\ \beta_q(p) &= \beta_q + \varrho_q F^{-1}(p) \\ \beta_c(p) &= \beta_c \end{aligned} \quad (5.28)$$

The constant coefficient β_0 by default always has a quantile profile so as to enforce location effects. Solution (5.28) to model (5.27) therefore motivates the investigation of sparsity detection for both the β_c and β_q vectors, and also whether the assumption of the error distribution effects this.

In particular, consider a homoskedastic design in which ϱ is populated by zeros such that $\beta(p) = \beta \forall p$ where $\beta = (\beta'_c, \beta'_q)'$. These offer baseline DGPs in which, as per (5.28), only the constant has a non-constant quantile function. Sparsity patterns considered are:

- $\beta_{sparse} = (1, 1.5, 1, 0.5, 0.33, 0.25, 0_{1 \times (K-4)})$

- $\beta_{block} = (1, 0.5_{1 \times \frac{K}{5}}, 0_{1 \times 2 \frac{K}{5}}, 0.5_{1 \times \frac{K}{5}})$

To investigate how well the BQR_{SAVS} and BQR_{BIC} detect sparsity specific to a given quantile, we consider additionally a heteroskedastic set of DGPs where β_q is zero only for some p . In particular, we let the constant coefficient be zero at the median and an additional covariate be zero only for $0.15 < p < 0.85$.⁸ This is equivalent to making the ϱ be quantile varying as well. In particular ϱ_q for the quantile specific sparsity is defined as:

$$\varrho_q^* = \varrho_q [I(u_t \leq F^{-1}(0.15)) + I(u_t \geq F^{-1}(0.85))] \quad (5.29)$$

Sparsity patterns that are considered for the quantile specific sparse Monte Carlo designs are:

- $\beta_{sparse,c} = (0, 0.5, 0.33, 0.25, 0_{1 \times (K-4)})$, $\beta_q = 0$ and $\beta_0 = 0$
- $\beta_{block,c} = (0, 0.5_{1 \times \frac{K}{5}}, 0_{1 \times 2 \frac{K}{5}}, 0.5_{1 \times \frac{K}{5}})$, $\beta_q = 0$ and $\beta_0 = 0$

For the homoskedastic DGPs, we consider standard normal and student-t-distributed errors with 3 degrees of freedom. For heteroskedastic DGPs, we also consider standard normal error and one further variant. Since it may also be plausible to find that there is sparsity across p , but with constant non-zero coefficients for non-sparse quantiles, we consider as a final heteroskedastic design the quantile varying coefficient to be -0.5 for $p \in (0, 0.15]$ and 0.5 for $p \in [0.85, 1)$. For clarity, the quantile

⁸ Centering the quantile function of β_0 on 0 helps identifying quantile profiles on other covariates on account of tails of coefficients being 'aligned'

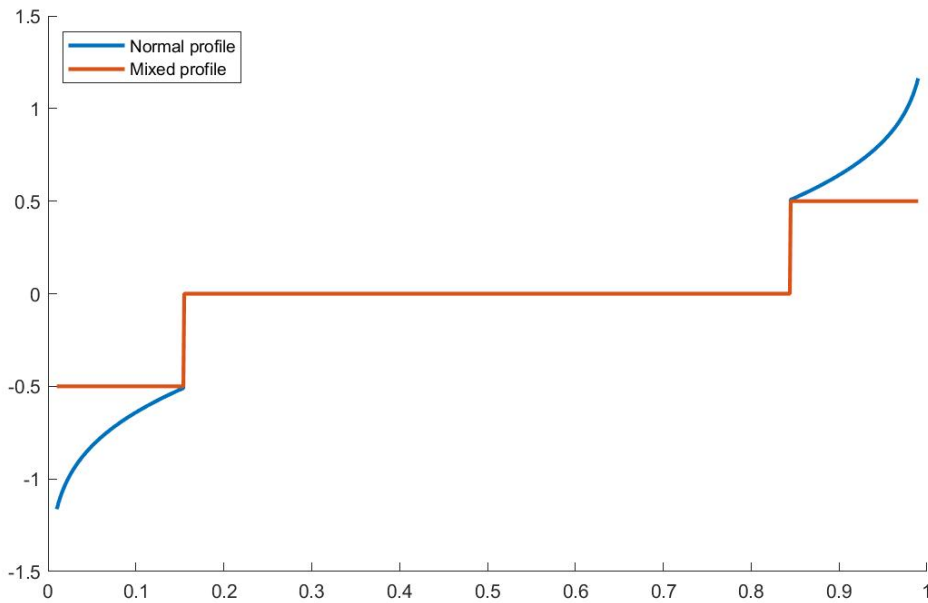


Figure 5.2: Quantile specific sparsity profiles

profiles for the quantile specific sparse coefficient are presented in figure (5.2). The quantile profile of the quantile varying beta coefficients for the 4 designs are summarised in table (5.1).

For all DGPs, we generate 50 synthetic data sets and set $K = 100$ (constant not included). The design matrix, X , is simulated using a multivariate Gaussian with zero mean and Toeplitz covariance structure, Ω , where $\Omega_{i,j} = \rho^{|i-j|}$ and $\rho = 0.5$. There are two sample sizes considered $T_i \in \{100, 500\}$.

Performance of the priors are gauged in terms of coefficient bias and measures of correct variable selection. Coefficient bias is calculated as the average (over Monte Carlo iterations) root mean deviation to the true quantile regression coefficients:

$$\frac{1}{iter} \sqrt{\|\hat{\beta}(p) - \beta(p)\|_2^2} \quad (5.30)$$

where $iter$ is the number of Monte Carlo iterations, and $\hat{\beta}(p)$ is the mean posterior vector of the estimator.

Variable selection will be evaluated with Matthew's correlation coefficient (MCC) and the hit rate. While the MCC gives an overall measure of goodness of variable selection based on the true and false positives and negatives, the hit rate tells us how well we identify the true positives. The MCC and the hit rate can be calculated from the confusion matrix with the following formulas:

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad (5.31)$$

$$Hit\ rate = \frac{TP}{TP + FN} \quad (5.32)$$

where T stands for "True", F stands for "False", N stands for "Negative", and P stands for "Positive". The MCC and hit rate are commonly used for binary classification problems but high-dimensional problems can be easily amended to fit the framework: non-zero coefficients have a value of 1 and zero coefficients have a value of 0. The same way coefficients shrunk to 0 receive a value of 0, while coefficients that are not 0 get a value of 1. This way a confusion matrix can be constructed from which the measures can be calculated. The MCC is between -1 (worst) and +1 (best), while the hit rate will range between 0 (worst) and 1 (best).

5.6.2 Monte Carlo Results

The coefficient bias results of the designs for a selection of quantiles are presented in table (5.2), the results for the MCC are shown in table (5.3), and the results of for the hit rates can be seen in table (5.4)⁹. The tables tell a very clear story: sparsification

⁹ Note that the results presented in tables (5.2)-(5.4) do not include the ALD adjustment term \bar{Z}_p .

Design	β profiles
y_1	$\beta_0(p) = \beta_0 + F_{N(0,1)}^{-1}(p)$
y_2	$\beta_0(p) = \beta_0 + F_{T(3)}^{-1}(p)$
y_3	$\beta_0(p) = \beta_0 + F_{N(0,1)}^{-1}(p)$ $\beta_1(p) = \beta_1 + [I(u_t \leq F_{N(0,1)}^{-1}(0.15)) + I(u_t \geq F_{N(0,1)}^{-1}(0.85))]F_{N(0,1)}^{-1}(p)$
y_4	$\beta_0(p) = \beta_0 + F_{N(0,1)}^{-1}(p)$ $\beta_1(p) = \beta_1 + [-\frac{1}{2}I(u_t \leq F_{N(0,1)}^{-1}(0.15)) + \frac{1}{2}I(u_t \geq F_{N(0,1)}^{-1}(0.85))]$

Table 5.1: Summary of simulation setups

via the BQR_{SAVS} and BQR_{BIC} improves or yields similar coefficient bias compared to the un-sparsified posteriors, while yielding more often than not better variable selection accuracy than the $SSVSBQR$, independent of sparsity, in-sample length and error-distribution. The BQR_{BIC} procedure stands out particularly in the sparse DGPs in which this sparsification procedure can reduce bias close to half for some quantiles in the larger sample setting compared to the un-sparsified posterior and offers further bias reduction compared to the BQR_{SAVS} . Results including the adjustment term \bar{Z}_p , presented in the appendix shows that the results are largely unaffected as expected.

Baseline DGPs

Focusing first on bias for the un-sparsified posterior, one can see from the first two columns of table (5.2), that the $HSBQR$ and $LBQR$ do similarly well in the sparse DGPs, while clearly outperforming the $SSVSBQR$. Although, all priors reduce bias with more data, the overall tendency doesn't change in the larger $T=500$ DGP. The largest difference among the SSVS, Lasso and horseshoe priors manifests itself in the block DGP's where the $HSBQR$ does consistently better, especially in the tails and in the smaller sample size of $T=100$. These findings corroborate the findings of Kohns and Szendrei (2020) that the adaptive shrinkage from global local priors are able to capture different sparsity blocks among the covariates and adapt to low data settings.

	0.05	0.25	0.50	0.75	0.95	0.05	0.25	0.50	0.75	0.95	0.05	0.25	0.50	0.75	0.95					
	y_1					y_2					y_3					y_4				
T=100																				
Sparse																				
<i>SSVSBQR</i>	0.206	0.210	0.217	0.212	0.221	0.227	0.209	0.195	0.201	0.238	0.235	0.241	0.227	0.253	0.243	0.193	0.219	0.223	0.222	0.194
<i>LBQR</i>	0.132	0.079	0.068	0.084	0.143	0.173	0.082	0.067	0.084	0.164	0.158	0.081	0.063	0.087	0.153	0.133	0.084	0.065	0.087	0.140
<i>LBQR_{s,AVS}</i>	0.132	0.076	0.065	0.082	0.143	0.173	0.080	0.065	0.082	0.164	0.138	0.082	0.061	0.088	0.154	0.133	0.086	0.062	0.087	0.141
<i>LBQR_{R,BC}</i>	0.132	0.076	0.065	0.082	0.143	0.173	0.080	0.065	0.081	0.164	0.158	0.081	0.064	0.088	0.153	0.133	0.085	0.062	0.086	0.141
<i>HSBQR</i>	0.119	0.094	0.085	0.097	0.127	0.172	0.099	0.087	0.097	0.176	0.165	0.104	0.075	0.100	0.157	0.131	0.099	0.077	0.099	0.131
<i>HSBQR_{s,AVS}</i>	0.108	0.074	0.064	0.080	0.117	0.164	0.080	0.066	0.081	0.169	0.157	0.095	0.061	0.090	0.148	0.123	0.089	0.061	0.088	0.124
<i>HSBQR_{R,BC}</i>	0.101	0.063	0.058	0.072	0.109	0.159	0.067	0.056	0.073	0.164	0.149	0.096	0.060	0.090	0.140	0.119	0.090	0.060	0.088	0.119
Block																				
<i>SSVSBQR</i>	0.320	0.325	0.325	0.336	0.336	0.339	0.333	0.322	0.328	0.360	0.338	0.371	0.362	0.359	0.353	0.337	0.340	0.353	0.339	0.332
<i>LBQR</i>	0.369	0.328	0.327	0.323	0.382	0.338	0.329	0.326	0.328	0.342	0.414	0.352	0.339	0.355	0.419	0.410	0.348	0.342	0.346	0.411
<i>LBQR_{s,AVS}</i>	0.368	0.326	0.325	0.321	0.381	0.337	0.327	0.324	0.326	0.341	0.413	0.349	0.336	0.353	0.418	0.410	0.346	0.339	0.344	0.411
<i>LBQR_{R,BC}</i>	0.368	0.326	0.326	0.321	0.381	0.337	0.327	0.324	0.326	0.341	0.413	0.349	0.336	0.352	0.418	0.410	0.346	0.339	0.344	0.410
<i>HSBQR</i>	0.250	0.248	0.251	0.253	0.244	0.260	0.240	0.244	0.249	0.272	0.288	0.293	0.290	0.286	0.278	0.258	0.267	0.269	0.263	0.255
<i>HSBQR_{s,AVS}</i>	0.252	0.244	0.244	0.247	0.243	0.263	0.236	0.239	0.244	0.271	0.289	0.287	0.282	0.280	0.277	0.261	0.263	0.262	0.258	0.256
<i>HSBQR_{R,BC}</i>	0.259	0.246	0.246	0.247	0.247	0.271	0.241	0.244	0.247	0.275	0.287	0.287	0.282	0.280	0.277	0.258	0.263	0.263	0.259	0.254
T=500																				
Sparse																				
<i>SSVSBQR</i>	0.133	0.095	0.087	0.095	0.131	0.165	0.066	0.083	0.095	0.170	0.159	0.084	0.069	0.087	0.156	0.134	0.088	0.073	0.083	0.132
<i>LBQR</i>	0.092	0.053	0.042	0.054	0.091	0.138	0.060	0.044	0.060	0.147	0.119	0.056	0.039	0.056	0.119	0.102	0.053	0.036	0.055	0.105
<i>LBQR_{s,AVS}</i>	0.092	0.046	0.033	0.049	0.090	0.139	0.053	0.035	0.056	0.147	0.119	0.056	0.029	0.056	0.119	0.102	0.052	0.027	0.055	0.105
<i>LBQR_{R,BC}</i>	0.092	0.041	0.029	0.046	0.090	0.138	0.048	0.032	0.053	0.147	0.119	0.056	0.031	0.056	0.119	0.101	0.051	0.030	0.054	0.105
<i>HSBQR</i>	0.099	0.072	0.062	0.074	0.103	0.152	0.074	0.061	0.078	0.153	0.137	0.066	0.047	0.068	0.134	0.110	0.065	0.047	0.067	0.113
<i>HSBQR_{s,AVS}</i>	0.086	0.051	0.038	0.054	0.089	0.145	0.057	0.037	0.060	0.145	0.125	0.057	0.025	0.059	0.122	0.096	0.056	0.026	0.058	0.101
<i>HSBQR_{R,BC}</i>	0.074	0.043	0.029	0.047	0.076	0.139	0.050	0.030	0.053	0.138	0.106	0.063	0.030	0.064	0.107	0.083	0.060	0.030	0.062	0.090
Block																				
<i>SSVSBQR</i>	0.150	0.101	0.094	0.104	0.145	0.177	0.101	0.091	0.101	0.180	0.190	0.097	0.079	0.098	0.180	0.162	0.097	0.084	0.099	0.156
<i>LBQR</i>	0.305	0.100	0.079	0.094	0.306	0.290	0.097	0.079	0.099	0.288	0.340	0.102	0.074	0.098	0.327	0.322	0.100	0.073	0.098	0.320
<i>LBQR_{s,AVS}</i>	0.304	0.101	0.080	0.096	0.305	0.289	0.099	0.081	0.100	0.287	0.340	0.104	0.074	0.100	0.326	0.321	0.103	0.073	0.100	0.319
<i>LBQR_{R,BC}</i>	0.304	0.100	0.079	0.094	0.305	0.289	0.097	0.079	0.099	0.287	0.340	0.102	0.073	0.098	0.326	0.321	0.100	0.073	0.098	0.319
<i>HSBQR</i>	0.142	0.104	0.094	0.105	0.143	0.173	0.103	0.091	0.104	0.175	0.189	0.099	0.083	0.097	0.184	0.155	0.100	0.084	0.099	0.154
<i>HSBQR_{s,AVS}</i>	0.137	0.092	0.082	0.093	0.137	0.172	0.092	0.078	0.093	0.171	0.186	0.090	0.072	0.089	0.182	0.151	0.091	0.073	0.091	0.150
<i>HSBQR_{R,BC}</i>	0.138	0.095	0.085	0.096	0.137	0.173	0.094	0.080	0.095	0.171	0.187	0.090	0.072	0.088	0.182	0.153	0.091	0.073	0.090	0.152

Table 5.2: Root Mean Coefficient Bias Results

	0.05	0.25	0.50	0.75	0.95	0.05	0.25	0.50	0.75	0.95	0.05	0.25	0.50	0.75	0.95					
	y_1					y_2					y_3					y_4				
T=100																				
Sparse																				
<i>SSVSBQR</i>	0.243	0.275	0.269	0.292	0.325	0.266	0.273	0.305	0.304	0.353	0.225	0.188	0.180	0.184	0.217	0.262	0.210	0.172	0.205	0.266
<i>LBQR_{sAVS}</i>	0.589	0.710	0.789	0.781	0.660	0.570	0.712	0.797	0.791	0.680	0.428	0.581	0.558	0.546	0.432	0.442	0.551	0.545	0.538	0.399
<i>LBQR_{R_{BIC}}</i>	0.589	0.720	0.829	0.814	0.656	0.574	0.717	0.824	0.821	0.680	0.433	0.586	0.518	0.528	0.436	0.439	0.548	0.572	0.548	0.404
<i>HSBQR_{sAVS}</i>	0.390	0.532	0.560	0.550	0.513	0.382	0.523	0.554	0.556	0.503	0.350	0.422	0.442	0.415	0.353	0.394	0.426	0.429	0.411	0.396
<i>HSBQR_{R_{BIC}}</i>	0.552	0.720	0.782	0.772	0.695	0.541	0.722	0.793	0.789	0.689	0.468	0.495	0.542	0.488	0.472	0.510	0.504	0.539	0.493	0.506
Block																				
<i>SSVSBQR</i>	0.420	0.386	0.381	0.384	0.405	0.416	0.375	0.392	0.382	0.423	0.374	0.319	0.314	0.330	0.394	0.410	0.352	0.311	0.376	0.417
<i>LBQR_{sAVS}</i>	0.255	0.429	0.488	0.473	0.261	0.252	0.437	0.496	0.471	0.256	0.230	0.415	0.439	0.410	0.217	0.232	0.424	0.427	0.426	0.222
<i>LBQR_{R_{BIC}}</i>	0.255	0.424	0.478	0.471	0.261	0.254	0.437	0.492	0.471	0.256	0.229	0.411	0.436	0.409	0.217	0.232	0.421	0.424	0.423	0.222
<i>HSBQR_{sAVS}</i>	0.387	0.408	0.421	0.410	0.393	0.406	0.430	0.440	0.420	0.401	0.337	0.342	0.343	0.351	0.344	0.383	0.391	0.385	0.395	0.383
<i>HSBQR_{R_{BIC}}</i>	0.423	0.442	0.459	0.452	0.436	0.436	0.458	0.471	0.462	0.444	0.251	0.327	0.350	0.333	0.255	0.276	0.352	0.369	0.355	0.272
T=500																				
Sparse																				
<i>SSVSBQR</i>	0.275	0.308	0.305	0.309	0.314	0.294	0.311	0.320	0.324	0.318	0.284	0.331	0.304	0.334	0.293	0.301	0.316	0.291	0.348	0.325
<i>LBQR_{sAVS}</i>	0.704	0.745	0.755	0.789	0.790	0.740	0.707	0.758	0.792	0.809	0.571	0.797	0.718	0.773	0.593	0.589	0.783	0.757	0.776	0.571
<i>LBQR_{R_{BIC}}</i>	0.708	0.903	0.924	0.926	0.805	0.745	0.878	0.907	0.924	0.820	0.594	0.836	0.835	0.839	0.603	0.598	0.848	0.852	0.823	0.590
<i>HSBQR_{sAVS}</i>	0.401	0.538	0.563	0.539	0.470	0.393	0.570	0.579	0.555	0.470	0.395	0.665	0.635	0.642	0.416	0.422	0.658	0.634	0.636	0.441
<i>HSBQR_{R_{BIC}}</i>	0.711	0.899	0.899	0.901	0.817	0.713	0.909	0.901	0.900	0.837	0.699	0.810	0.847	0.805	0.722	0.724	0.819	0.845	0.797	0.747
Block																				
<i>SSVSBQR</i>	0.682	0.725	0.711	0.713	0.696	0.694	0.730	0.722	0.726	0.708	0.638	0.753	0.768	0.768	0.657	0.679	0.760	0.744	0.753	0.707
<i>LBQR_{sAVS}</i>	0.584	0.937	0.959	0.959	0.595	0.595	0.951	0.963	0.955	0.591	0.552	0.935	0.956	0.945	0.549	0.556	0.941	0.958	0.937	0.563
<i>LBQR_{R_{BIC}}</i>	0.584	0.930	0.956	0.950	0.595	0.595	0.937	0.957	0.946	0.591	0.552	0.925	0.952	0.938	0.549	0.556	0.935	0.950	0.930	0.563
<i>HSBQR_{sAVS}</i>	0.691	0.800	0.810	0.799	0.696	0.708	0.804	0.818	0.812	0.712	0.612	0.830	0.863	0.847	0.616	0.679	0.823	0.855	0.843	0.677
<i>HSBQR_{R_{BIC}}</i>	0.745	0.929	0.947	0.930	0.766	0.749	0.934	0.953	0.941	0.783	0.433	0.867	0.906	0.880	0.441	0.439	0.833	0.887	0.847	0.449

Table 5.3: MCC Results

	0.05	0.25	0.50	0.75	0.95	0.05	0.25	0.50	0.75	0.95	0.05	0.25	0.50	0.75	0.95
T=100															
Sparse															
<i>SSVSBQR</i>	0.622	0.804	0.854	0.848	0.818	0.637	0.796	0.867	0.862	0.831	0.573	0.650	0.688	0.676	0.570
<i>LBQR_SAVS</i>	0.399	0.647	0.788	0.742	0.512	0.362	0.646	0.772	0.768	0.524	0.226	0.445	0.501	0.424	0.233
<i>LBQR_RBIC</i>	0.400	0.568	0.759	0.712	0.509	0.364	0.597	0.747	0.755	0.524	0.227	0.410	0.333	0.353	0.233
<i>HSBQR_SAVS</i>	0.576	0.814	0.844	0.835	0.769	0.567	0.792	0.827	0.822	0.759	0.530	0.589	0.611	0.573	0.550
<i>HSBQR_RBIC</i>	0.483	0.638	0.706	0.706	0.700	0.505	0.645	0.727	0.725	0.695	0.437	0.367	0.385	0.353	0.449
Block															
<i>SSVSBQR</i>	0.575	0.652	0.682	0.666	0.593	0.570	0.650	0.699	0.669	0.601	0.557	0.620	0.638	0.634	0.557
<i>LBQR_SAVS</i>	0.120	0.325	0.387	0.372	0.129	0.124	0.333	0.406	0.369	0.122	0.113	0.332	0.370	0.329	0.105
<i>LBQR_RBIC</i>	0.121	0.318	0.371	0.366	0.129	0.123	0.332	0.400	0.368	0.122	0.112	0.326	0.359	0.324	0.105
<i>HSBQR_SAVS</i>	0.554	0.651	0.686	0.672	0.583	0.569	0.665	0.692	0.671	0.585	0.548	0.631	0.649	0.639	0.555
<i>HSBQR_RBIC</i>	0.362	0.414	0.443	0.439	0.389	0.369	0.423	0.445	0.436	0.387	0.686	0.654	0.632	0.664	0.694
T=500															
Sparse															
<i>SSVSBQR</i>	0.875	0.973	0.984	0.973	0.964	0.905	0.980	0.989	0.984	0.949	0.857	0.963	0.990	0.976	0.874
<i>LBQR_SAVS</i>	0.569	0.919	0.950	0.938	0.698	0.608	0.880	0.937	0.929	0.712	0.413	0.886	0.918	0.869	0.442
<i>LBQR_RBIC</i>	0.573	0.851	0.886	0.896	0.700	0.611	0.819	0.874	0.888	0.712	0.410	0.848	0.733	0.838	0.428
<i>HSBQR_SAVS</i>	0.773	0.950	0.962	0.952	0.926	0.752	0.959	0.958	0.955	0.917	0.824	0.927	0.945	0.908	0.840
<i>HSBQR_RBIC</i>	0.637	0.845	0.844	0.855	0.813	0.648	0.864	0.851	0.855	0.827	0.648	0.752	0.755	0.746	0.664
Block															
<i>SSVSBQR</i>	0.969	1.000	1.000	0.999	0.983	0.966	0.999	1.000	0.999	0.980	0.938	0.991	0.998	0.992	0.938
<i>LBQR_SAVS</i>	0.500	0.974	0.996	0.986	0.499	0.509	0.980	0.993	0.984	0.500	0.473	0.963	0.991	0.964	0.476
<i>LBQR_RBIC</i>	0.500	0.975	0.996	0.987	0.499	0.509	0.981	0.993	0.985	0.500	0.474	0.965	0.990	0.968	0.476
<i>HSBQR_SAVS</i>	0.964	0.999	0.999	0.997	0.971	0.964	0.999	0.999	0.998	0.972	0.932	0.986	0.995	0.989	0.927
<i>HSBQR_RBIC</i>	0.951	0.986	0.991	0.985	0.956	0.955	0.989	0.993	0.990	0.954	0.968	0.984	0.994	0.986	0.965

Table 5.4: Hit Rate Results

With sparsification, one sees generally that the BQR_{SAVS} and BQR_{BIC} reduce bias nearly uniformly across DGPs. However, while the $LBQR$ and $HSBQR$ had similar bias in sparse DGPs, the $HSBQR_{BIC}$ has the lowest bias among all models, while trading blows with the $HSBQR_{SAVS}$ in the block DGPs. The $HSBQR$ coming ahead with sparsification indicates that signals are better identified through the horseshoe than the lasso prior.

The variable selection results in the first two columns of (5.3) and (5.4), mirror the bias results, while highlighting that the reason for the better performance from the BQR_{BIC} models comes from lower false positive rates. This can be deduced from the fact that while the hit rates in table (5.4) are higher for the other models, the MCC are lower. This also holds for the SSVS prior. Hence, both the $SSVSBQR$ and BQR_{SAVS} models tend to select too many noise variables.

Quantile specific Sparse DGPs

Similar to the baseline DGPs, the bias tendencies carry mostly also over to the quantile specific simulations, with the difference that the $HSBQR$ does exceptionally well in the tails of the block DGPs. This can be understood from the fact that adaptive shrinkage plays an even larger role with quantile specific sparsity, particularly in the more extreme tails which are data sparse (Koenker, 2005).

Adding sparsification, the BQR_{BIC} does better than BQR_{SAVS} in sparse DGPs, however, similarly to the baseline results, does not offer a systematic improvement over the BQR_{SAVS} in the block DGPs. This makes sense in the high-dimensional settings considered where capturing the quantile profile of the single covariate has a larger impact on fit in the sparse DGP than the block DGP, where many more covariates have non-zero coefficients.

In terms of variable selection, we find that the continuous priors considered outperform the $SSVSBQR$ in sparse DGPs and the $HSBQR_{BIC}$ does especially well in the tails, consistently having the highest MCC. The results in the block design deviate from the baseline results: here we see that the $SSVSBQR$ and BQR_{SAVS} models tend to do better than the $HSBQR_{BIC}$ in terms of the MCC, yet have similar or higher bias. Considering the hit rates, this difference indicates that the BQR_{BIC} now offers lower penalisation in order to capture the quantile effect from the quantile specific variable which the tick-loss function heavily tilts the predicted quantile towards. Allowing for proper ALD correction by including \bar{Z}_p as in (5.21) for these DGPs, however, allows the $HSBQR_{BIC}$ to again outperform in the tails or come close to the best performer in terms of variable selection. Looking at the hit rates, it is suggestive that the \bar{Z}_p correction implies some additional penalisation particularly in the tails which increases sparsity and decreases the amount of false positives from the sparse parts. Nevertheless, in terms of bias, the correction induces very small differences and vanishes with more data.

All in all, the simulations have shown that sparsification often lowers bias noticeably, or at worst holds bias stable, while improving variable selection compared to the SSVS. The BQR_{BIC} outperforms the BQR_{SAVS} , especially in sparse designs and in the tails where the extra flexibility of determining the penalisation from the data allows to capture the quantile profile. The results have also shown that although sparsification often helps the priors' performance, care should be taken in selecting the shrinkage prior a-priori.

5.7 Growth at Risk application

The initial motivation for growth-at-risk, as introduced by Adrian et al. (2019), is to measure the downside risks, or vulnerabilities, to real GDP growth, which are captured as the response of forecasted conditional quantiles to changes in summary measures of financial conditions. Such models provide useful information for policymakers, as it has been empirically and theoretically shown that crises stemming from the financial sector carry the risk of creating negative feedback loops between the real economy and the financial system (Jordà et al., 2015). However, as seen from recent macroeconomic developments, vulnerabilities to GDP growth may stem from various, possibly a priori unknown parts of the economy which entails that a larger information set is needed. Recent contributions to the GaR literature indeed conduct real time nowcasting exercises (Carriero et al., 2020b; Ferrara et al., 2020) and forecast density combination (Kohns and Szendrei, 2020; Korobilis, 2017) using larger macro data sets in combination with shrinkage priors in order to deal with parameter proliferation.

However, shrinkage priors, such as the very popular global-local priors, are continuous and are therefore hard to interpret in higher dimensions. For this application, similar to Kohns and Szendrei (2020), we are interested in creating quantile and combined density forecasts for real GDP growth based on the entire McCracken and Ng (2020) data base. The logic of combining forecasts of multiple quantiles is to allow for the full heterogeneity of the data's effects to permeate to the conditional forecast distribution, which has been shown in the previous literature to often outperform models focused on only location effects (Korobilis, 2017)¹⁰. However, unlike previous contributions, we apply the introduced sparsification methods to understand which variables drive the forecasting results in an effort to better communicate which subsets of the data set explain particular parts of the forecast distribution.

The McCracken and Ng (2020) database¹¹, including real GDP, consists of 248

¹⁰ This step should also be natural from the standpoint that conditional quantiles characterise empirical CDFs which in turn provide direct links to the latent forecasted probability density function.

¹¹ <https://research.stlouisfed.org/econ/mccracken/fred-databases/>

macroeconomic time series, starting from 1959Q1 at a quarterly frequency and is updated monthly. We take the quarter-on-quarter growth rate of annualized real GDP as our dependent variable and all others as independent covariates. These variables include a wide variety of macroeconomic effects which cover real, financial as well as national accounts data. Since not all series start at 1959Q1, for the growth at risk application, only variables that are available from 1970Q1 onwards were considered which gives 229 explanatory variables.

Predictions for our quantile methodology are generated from a direct forecasting model:

$$y_{t+h} = x_t' \beta + \epsilon_{t+h} \quad (5.33)$$

for $t = 1, \dots, T - h$, where h refers to the forecast horizon. We consider one- to four-quarter ahead forecast horizons ($h = 1, \dots, 4$). Following Yu and Moyeed (2001), the quantile specific predictive distribution can be obtained by marginalising out the uncertainty of the quantile regression parameter posterior $p(\beta_p | \bullet)$:

$$p(y_{t+h,p} | y_{1:t}) = \int p(y_{1:t} | \beta_p) p(\beta_p | y_{1:t}) d\beta_p, \quad (5.34)$$

where in (5.34) $y_{1:t}$ refers to in-sample observations for (y_1, \dots, y_t) . (5.34) is conveniently approximated through Monte Carlo integration, as posterior draws β_p^s are available from MCMC algorithms detailed in the appendix:

$$p(y_{t+h,p} | y_{1:t}) \approx \frac{1}{S} \sum_{s=1}^S p(y_{1:t} | \beta_p^s). \quad (5.35)$$

Note that for each MCMC iteration $p(\beta_p | \bullet)$, β_p^s , we obtain a sparsified vector through the BQR_{SAVS} and BQR_{BIC} procedure respectively, which are used again in (5.35) to generate separate predictive distributions. Predictive distributions are estimated for plain, BQR_{SAVS} sparsified and BQR_{BIC} sparsified posteriors of the $HSBQR$ and $LBQR$ as described in section (5.5) and the $SSVSBQR$ serves as a benchmark for the large data models. To further contrast our results to the well established 'vulnerable growth' model of Adrian et al. (2019), we estimate a Bayesian quantile regression with a fairly uninformative standard normal prior on the regression coefficients of only lagged real GDP and the NFCI index. This model is abbreviated in the tables below as ABG_{BQR} . Finally, to obtain a combined predictive density, we follow Gaglianone and Lima (2012) and Korobilis (2017) by estimating quantile predictive distributions for 19 equidistant quantiles, $p \in [0.05, 0.10, \dots, 0.90, 0.95]$, sort the stacked $vec(S \times 19)$ vector, and smooth this stacked vector via a Gaussian kernel¹². Although stacking quantile predictive distributions will inevitably create some quantile crossing, the combination of multiple neighbouring quantiles helps in providing probability density in tail regions, which,

¹² This can be implemented for example via the "kdensity" function in Matlab

due to the tick-loss function in the likelihood, are statistically data sparse relative to the median (Koenker, 2005).

To provide direct reference also to the original work of Adrian et al. (2019), we construct quantile and density forecasts based on their two-step estimation procedure where conditional quantiles are smoothed via the skew-t-distribution (Azzalini and Capitanio, 2003).¹³ Forecasts for all models are computed on an expanding window basis where the initial in-sample period uses the first 50 observations of the sample, which makes for 149-h forecast windows.¹⁴

To evaluate forecasts, we employ both measures of overall density accuracy and quantile specific fit. For the overall density, we consider two strictly proper scoring rules in the sense of Gneiting and Raftery (2007), the average log-predictive density score (LPDS) and average cumulative rank probability score (CRPS). Define T_0 as the initial in-sample period, taken to be (1970Q1-1982Q2), and T as the final forecast period (Q42019), then the LPDS and CRPS are calculated as follows:

$$\begin{aligned}
 LPDS &= \frac{1}{T - T_0 - h + 1} \sum_{t=1}^{T-T_0-h+1} \log\left(\frac{1}{S} \sum_{s=1}^S p(y_{T_0+t+h-1}|y_{1:T_0+t})\right) \\
 CRPS &= \frac{1}{T - T_0 - h + 1} \sum_{t=1}^{T-T_0-h+1} |y_{T_0+t+h-1} - \hat{y}_{T_0+t+h-1}^A| - \frac{1}{2} |\hat{y}_{T_0+t+h-1}^A - \hat{y}_{T_0+t+h-1}^B|,
 \end{aligned}
 \tag{5.36}$$

where $\hat{y}_{T_0+t+h-1}^A$ and $\hat{y}_{T_0+t+h-1}^B$ are two independent draws from the forecast distribution. To facilitate the discussion, the objective is to maximise the LPDS and minimise the CRPS. Since the LPDS and CRPS do not elucidate forecast calibration across different parts of the overall forecast distribution, we further make use of Rossi and Sekhposyan (2019)'s calibration test. If a model is well calibrated, then the forecasted CDF of the probability integral transforms (PITs), $g_{t+h} = \int_0^{y_{t+h}} p(u|y_{t+h})du$, should be statistically indistinguishable from a 45 degree line (Diebold et al., 1998). Following Rossi and Sekhposyan (2019), we report confidence bands around the 45 degree line to account for sampling uncertainty ¹⁵.

Lastly, as a measure of local fit, we make use of the quantile weighted CRPS (qwCRPS) to assess tail forecast performance. This metric is based on the quantile score (QS), as proposed by Gneiting and Ranjan (2011) and is calculated as follows: let $\hat{y}_{t+h,p}$ denote the expected quantile for real GDP growth h steps ahead, then the

¹³ Results for the skew-t big data models are reported in the appendix

¹⁴ The ABG model contains 135-h rolling forecasts due to the availability of the NFCI index.

¹⁵ The confidence bands of Rossi and Sekhposyan (2019) should be taken as general guidance, as strictly speaking, they are derived using a rolling window of estimation, while we use an expanding window. As recommended by Rossi and Sekhposyan (2019), uncertainty intervals for higher order forecasts ($h > 1$), are computed using bootstrapping.

QS is computed as:

$$QS_{t+h,p} = (y_{t+h} - \hat{y}_{t+h,p}) \times (p - \mathbb{I}(y_{t+h} - \hat{y}_{t+h,p})). \quad (5.37)$$

The qwCRPS is thus given by:

$$qwCRPS_{t+h} = \int_0^1 w_p QS_{t+h,p} dp, \quad (5.38)$$

where w_p denotes a weighting scheme to evaluate specific parts of the forecast density. Since quantile regression is often used to capture downside risks, we use $w_p = (1-p)^2$, as suggested by Gneiting and Ranjan (2011) for left tail evaluation.

5.7.1 Forecast Density Evaluation Results

Point and density scoring results for all forecast horizons are reported in table (5.5), where stars indicate statistical significance as per the Diebold et al. (1998) test and calibration results are graphed in figure (5.4). The density scores show generally that there is a clear benefit to big data, as average LPDS are often several order of magnitudes higher, and CRPS significantly lower, especially for the *HSBQR* compared to either of the ABG models. The cumulative log-scores over time in graph (5.3) highlight that the vulnerable growth models underperform particularly during the financial crisis where the realisations of GDP growth seem to fall far into the tails of their forecast distributions, thereby being heavily penalised^{16 17}. In line with the simulation results, we find that the *HSBQR* outperforms both the *SSVSBQR* and *LBQR*, and that the *BQR_{BIC}* is clearly preferred for sparsification compared to the *BQR_{SAVS}*, where *BQR_{BIC}* sparsification often increases performance, as gauged by the CRPS, or only slightly worsens it as per the LPDS. Also similar to the simulations, the quality of sparsification is again dependent on the quality of shrinkage. This can be seen from the generally worse density scores of the *LBQR* models. Interestingly, for all models considered, forecast accuracy increases with the forecast horizon. This may stem the fact that changes in parts of the economy only affect aggregate output with a lag. Similar to much of the previous macro forecasting literature (Stock and Watson, 1998; Banerjee and Marcellino, 2006; Edge et al., 2010) we also find that simpler autoregressive models such as the ABG remain

¹⁶ The large performance discrepancy compared to the CRPS, therefore makes sense given that the LPDS penalises low mass event very heavily.

¹⁷ Fitting the skew-t distribution to the ABG model, markedly improves overall density scoring results and calibration at the one-quarter-ahead forecast horizon, where gains are largest for the middle to left parts of the density, as seen in figure (5.4). This gives an indication that the low-dimensional information set cannot provide enough information between quantiles, for which the skew-t provides extra spread. This in particular aids log-score penalisation. We find that applying the skew-t to the big-data models, doesn't not change performance qualitatively vis-a-vis ABG models. The table with the skew-t results are provided in the appendix.

	MSFE	LPDS	CRPS	qwCRPS	MSFE	LPDS	CRPS	qwCRPS
	$h=1$				$h=2$			
ABG_{BQR}	0.504	-11.607	0.313	1.049	0.404	-7.870	0.254	0.891
$ABG_{BQR-Skt}$	0.519	-1.727	0.301	1.026	0.426	-4.102	0.277	0.970
$SSVSBQR$	0.660***	-0.955***	0.337***	1.492	0.433***	-0.622**	0.235***	1.106
$HSBQR$	0.533*	-0.858***	0.296***	0.988*	0.411***	-0.558**	0.222***	0.804*
$HSBQR_{BIC}$	0.519	-0.895***	0.291***	0.948**	0.422***	-1.064**	0.227***	0.794*
$HSBQR_{SAVS}$	0.536	-0.976***	0.296***	0.968	0.468**	-2.856*	0.254***	0.855
$LBQR$	0.551*	-4.603**	0.305***	1.091	0.441***	-2.522*	0.238***	0.863
$LBQR_{BIC}$	0.552	-4.628**	0.305***	1.088	0.441***	-2.517*	0.238***	0.863
$LBQR_{SAVS}$	0.572	-4.832***	0.316***	1.085	0.467***	-3.006*	0.250***	0.859
	$h=3$				$h=4$			
ABG_{BQR}	0.378	-6.825	0.237	0.847	0.351	-4.387	0.235	0.844
$ABG_{BQR-Skt}$	0.397	-1.006	0.249	0.857	0.367	-1.252	0.251	0.875
$SSVSBQR$	0.391***	-0.797	0.219***	1.027	0.394***	-0.796	0.221***	1.038
$HSBQR$	0.390***	-0.649	0.215***	0.833	0.376**	-0.657	0.213***	0.773
$HSBQR_{BIC}$	0.397***	-1.589	0.219***	0.817	0.374**	-0.933	0.213***	0.761
$HSBQR_{SAVS}$	0.435*	-2.913	0.237***	0.830	0.391	-3.605	0.216***	0.757
$LBQR$	0.422***	-2.820	0.230***	0.872	0.382***	-3.498	0.216***	0.809
$LBQR_{BIC}$	0.423***	-2.920	0.230***	0.870	0.383***	-3.603	0.216***	0.809
$LBQR_{SAVS}$	0.442**	-4.323	0.239***	0.860	0.409**	-4.949	0.230***	0.823

Table 5.5: Forecast Evaluation Results

competitive in point forecasts, but fair far worse in density forecasts.

In terms of capturing downside risks, the qwCRPS results show that the $HSBQR_{BIC}$ consistently produces the best left tail performance, and significantly outperforms the ABG models at one-quarter and two-quarters ahead. The BQR_{BIC} sparsification procedure in fact increases tail performance both for the $HSBQR$ and $LBQR$.

Calibration based on PITs are able to shed more light on the differing performances. Figure (5.4) shows the models' PIT CDFs with confidence intervals computed as suggested by Rossi and Sekhposyan (2019). Corroborating the good density scores of the $HSBQR$ variants and the $SSVSBQR$, figure (5.4) shows that these models are also well calibrated, especially in the left tail at all horizons. The ABG

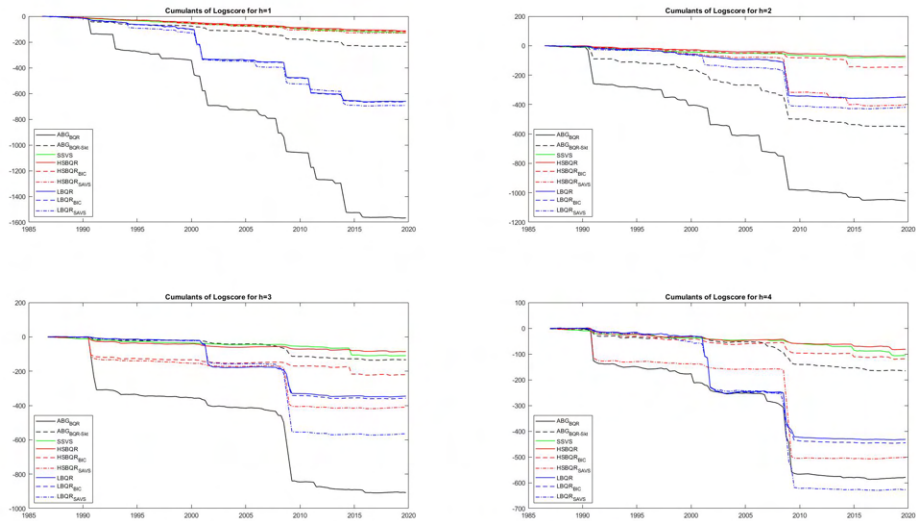


Figure 5.3: Cumulative log-scores.

models in turn tend to underestimate the lower tail which helps explain the bad log-score performance during the financial crisis in figure (5.3). Among the big-data models, the general tendency from the calibration results is that the $SSVSBQR$ slightly overpredicts the left tail and underpredicts the right tail, while the $LBQR$ tends to underpredict both tails. The $HSBQR$ in turn falls somewhere in the middle. In line with the density scores, sparsification via the BQR_{BIC} is able to preserve calibration, while the BQR_{SAVS} tends to overshrink. This is especially prevalent in the left to middle quantiles of the $HSBQR_{SAVS}$ at one-year ahead predictions, where more observations fall below those quantiles.

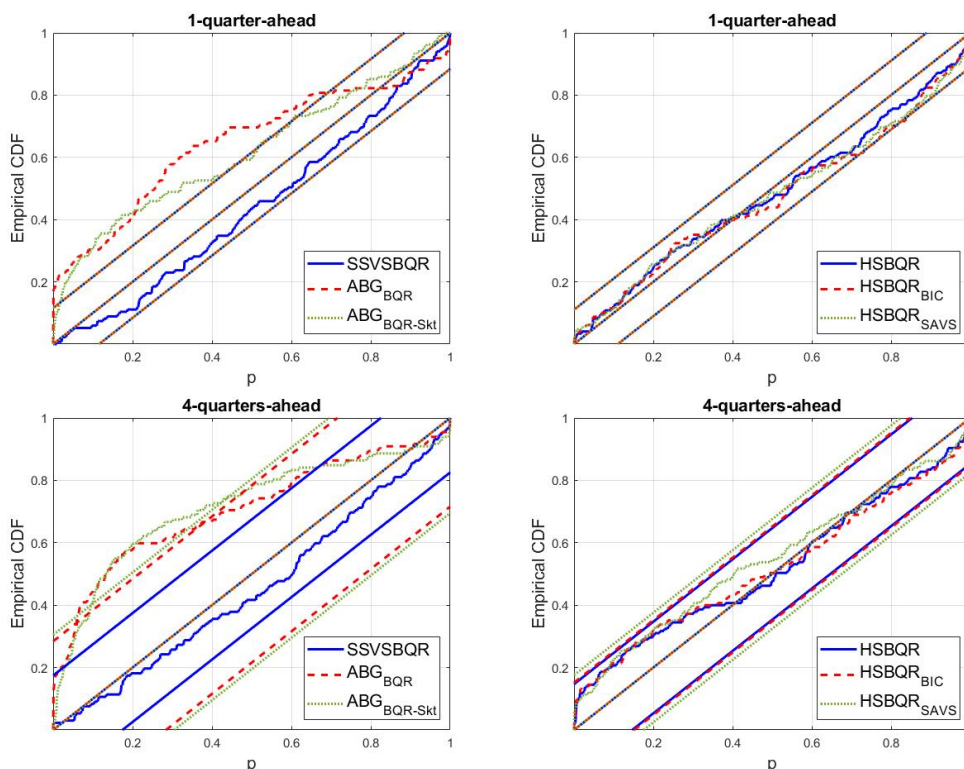


Figure 5.4: Rossi and Sekhposyan (2019) calibration test.

5.7.2 Quantile Specific Variable Inclusion

The results from the density and calibration tests have shown that the BQR_{BIC} variable selection procedure is able to preserve the fit of the un-sparsified forecast distributions, and even improve left tail accuracy, which gives credibility to the fact that most salient features of the models have been captured despite sparsification. In order to analyse variable importance both over time and across quantiles, we provide in figures (5.5), (5.6), (5.7), and (5.9) heatmaps of variable inclusion probability for each prior and sparsification variant. Since there are over 200 variables, showing all variables on the axis in one figure is impossible. As such, we will only show the highest inclusion probability among variables in the a given data group. Since there are 14 groups, this makes the figures much easier to parse and interpret. The choice

of using the maximum inclusion probability instead of the average inclusion probability, is that certain variables are selected very frequently in the group, while other variables are rarely chosen. As such, using the average inclusion probability per group, would hide how frequently these key variables are selected. Nevertheless, for completeness, the full heatmaps are shown in the appendix. While figure (5.5), (5.6), and (5.7) give a general static picture of variable selection across quantiles¹⁸, figure (5.9) shows variable inclusion probabilities for left tail (maximum inclusion probability of the average of 5th-15th quantiles), right tail (maximum inclusion probability of the average of 85th to 95th quantiles) as well as the middle regions (maximum inclusion probability of the average of 45th-55th quantiles) across time to showcase time variation in variable selection. Since it is of interest to understand differences in variable selection compared to widely used conditional mean models, we present at the bottom of figures (5.5) and (5.7) variable inclusion probabilities with each of the priors considered for the Bayesian quantile regression, however applied to a normal observational likelihood¹⁹. To make the graphs more easily interpretable, we present the covariates on the horizontal axis in terms of general variable groupings given by the FRED-QD data base. For example, series in the NIPA group refer to "National Income and Product Account" information that spans information such as aggregate consumption, investment and government expenditures. Exact variable definitions for each group can be found in the appendix of McCracken and Ng (2020).

Variable Inclusion Across Quantiles

For all quantile models under consideration, figures (5.5) - (5.7) clearly show that there is substantial quantile varying sparsity. For instance, one can see for the $HSBQR_{BIC}$ model in figure (5.5) at one-quarter ahead, that "Employment and Unemployment" data and "Household Balance Sheets" information are more important for the right tail, while playing only a marginal role in the left tail. At one-year ahead forecasts, variable importance shifts in the left tail toward "Interest Rates" and "Consumer Sentiment", while the right tail has higher inclusion probability for "Money and Credit" and "Non-Household Balance Sheets". The former information can be interpreted as important sectors that predict both the 2001 as well as great financial crisis. As expected, middle quantiles draw information from either tails, displaying variable inclusion patterns which overlap both with the left and right tail.²⁰

A commonality among, especially, the one-quarter ahead forecasts is the high

¹⁸ inclusion probabilities are first average over time per quantile, and then the max inclusion probability is used per data group

¹⁹ Details on the conditional mean models for the different priors is provided in the appendix.

²⁰ Note that variable groups overlapping need not signify that the fitted quantiles contain the same information since coefficients may vary across quantiles.

inclusion probability of NIPA variables at all quantiles. Since the NIPA variable group contains national account information which gets aggregated via the national account identity to GDP, it is natural to interpret this variable category as capturing information inherent in lagged GDP. In fact taking the first principal component of the four most included NIPA variables by the $HSBQR_{BIC}$ yields a series that is highly correlated with GDP growth. The fact that these series are included for the majority of the quantiles presents a departure from the previous literature such as Adrian et al. (2019) who find that current economic conditions, as captured by lagged GDP only forecast the median of the conditional distribution.

Contrasting these results to conditional mean models, we find that these tend to select variable groups which have high inclusion probability not only in the median, but also the tails. To underscore this, figure (5.8) plots the correlation coefficient between left, right and middle quantiles of the $HSBQR_{BIC}$ in a rolling 30 quarter window with point forecasts of the conditional mean horseshoe prior model with SAVS sparsification. While this figure is not a conclusive test of which quantiles influence the mean the most, it does highlight that correlation is highest with the median and right tail during tranquil times, while in recession periods, as indicated by grey bars, correlation with the lower quantiles increases, hence signifying that information spills over from the left tail to the mean. This figure, also highlights that there are larger differences in correlation between the upper, middle and lower quantiles in the one-year ahead forecasts after the 2001 crisis compared to the one-quarter ahead forecasts. This may reflect the fact that interest rate and credit information propagate only with a lag to aggregated economic activity (Romer, 2018) and therefore are increasingly picked up in higher order forecasts after periods in which monetary policy had changed a lot to counteract the recession. Lastly, in line with the recent macroeconomic forecasting literature such as Giannone et al. (2021), we also find that there is evidence for model uncertainty which can be deduced from the amount of yellow shaded areas at all forecasting horizons.

Comparing variable selection among priors and sparsification techniques, some further interesting patterns emerge. First, the BQR_{SAVS} algorithm with fixed penalisation induces much sparser models than under the more adaptive BQR_{BIC} penalisation for both continuous shrinkage priors, and in doing so, eliminates near all model uncertainty, as seen by the large white spaces in figure (5.6). Notably though, variables which are identified as strong signals by the $HSBQR_{BIC}$, tend to also appear with higher inclusion probability in the $HSBQR_{SAVS}$. As seen in the previous section, this heavy sparsification, however, comes at a large penalty on density scores and calibration, hurting particularly tail forecasting performance, as underscored in figure (5.3) around the financial crisis. Second, in line with the simulation results, we see also in the empirical application that the $LBQR$ tends in general to exert stronger shrinkage than the $HSBQR$, resulting in far sparser models

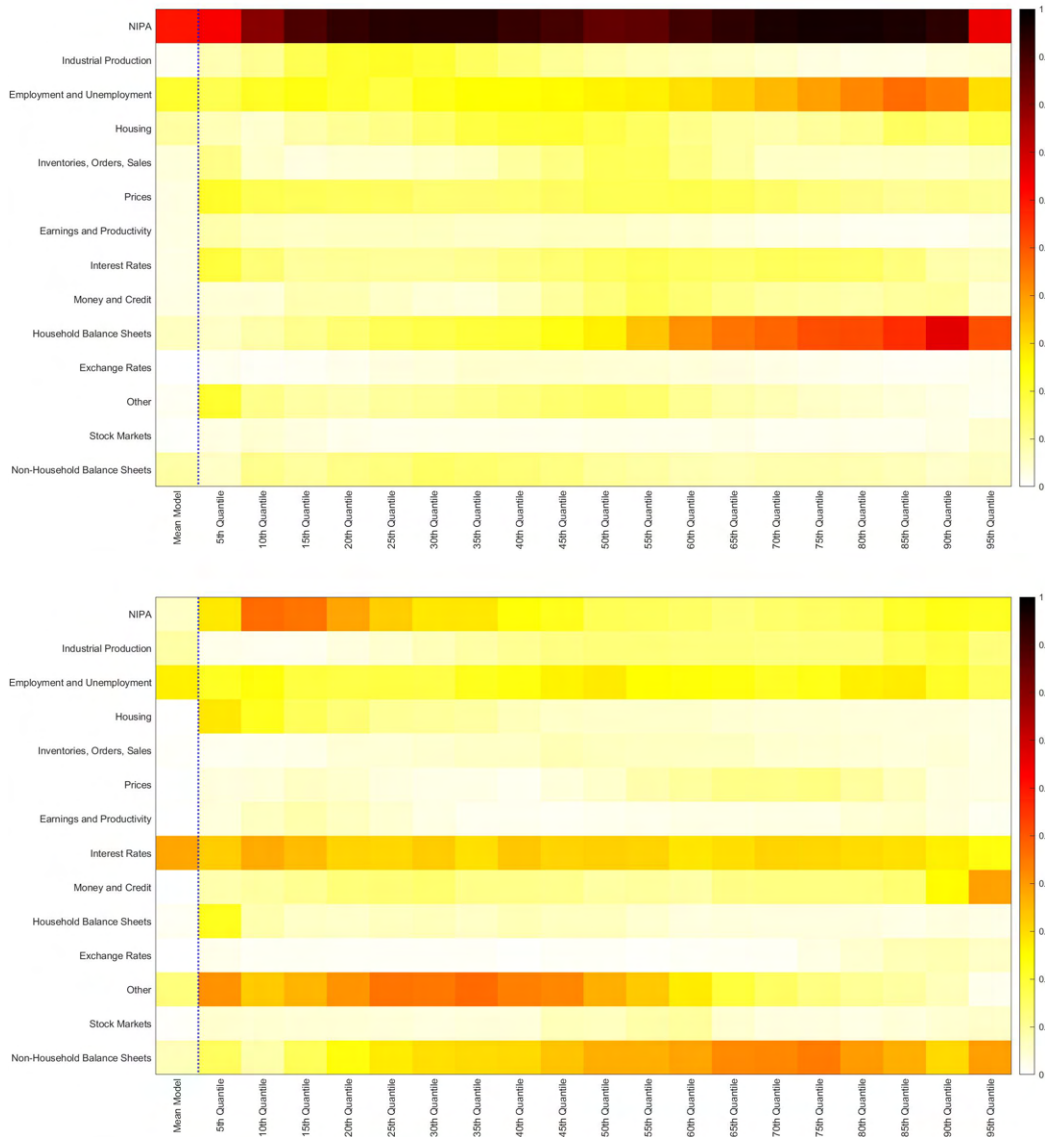


Figure 5.5: Maximum (per data group) of average inclusion probability for each quantile of the $HSBQR_{BIC}$ for $h=1$ (top) and $h=4$ (bottom). HS with SAVS for mean model included for reference

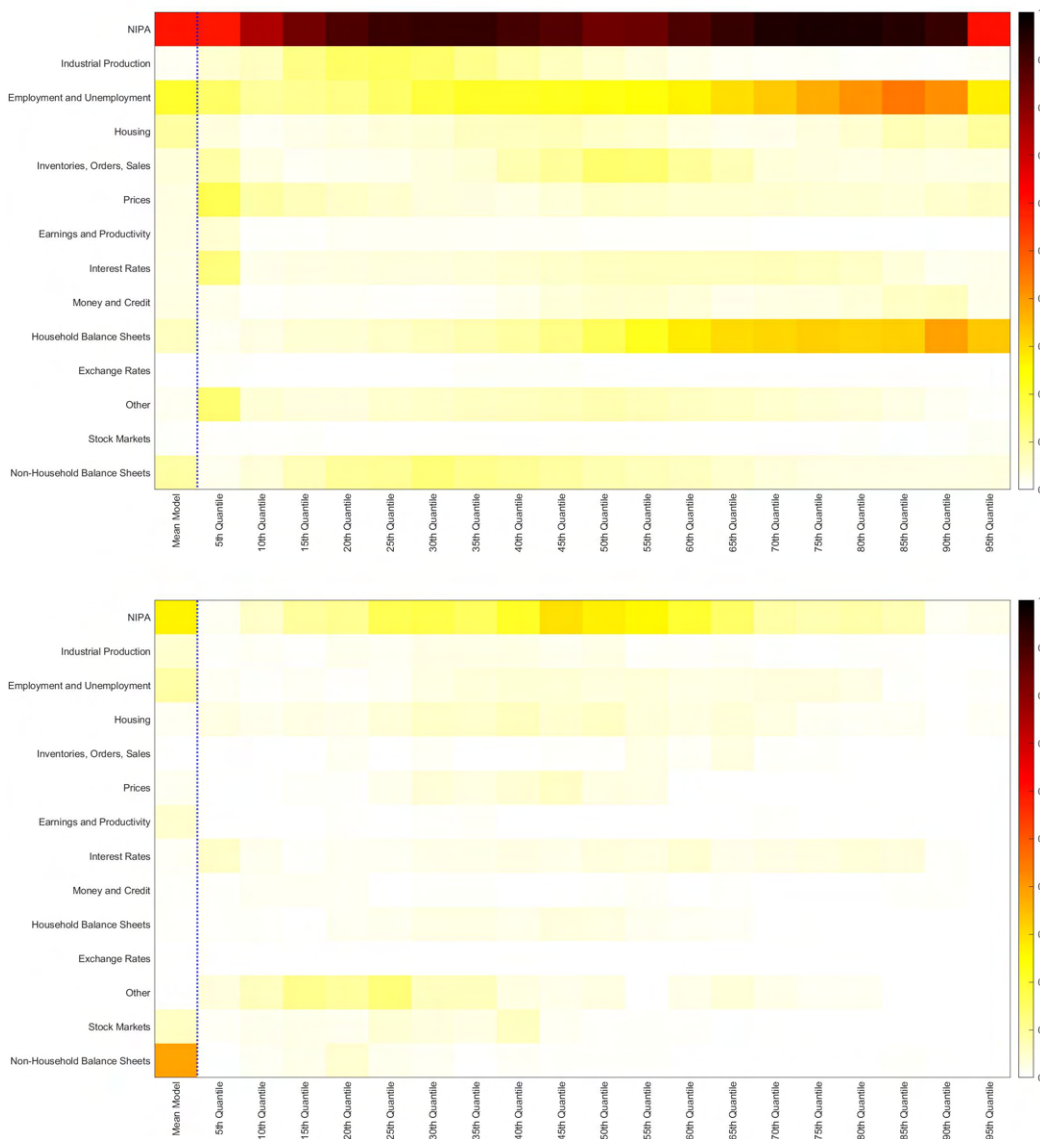


Figure 5.6: Maximum (per data group) of average inclusion probability for each quantile of the $HSBQR_{SAVS}$ (top) and the $LBQR_{SAVS}$ (bottom) for $h=1$

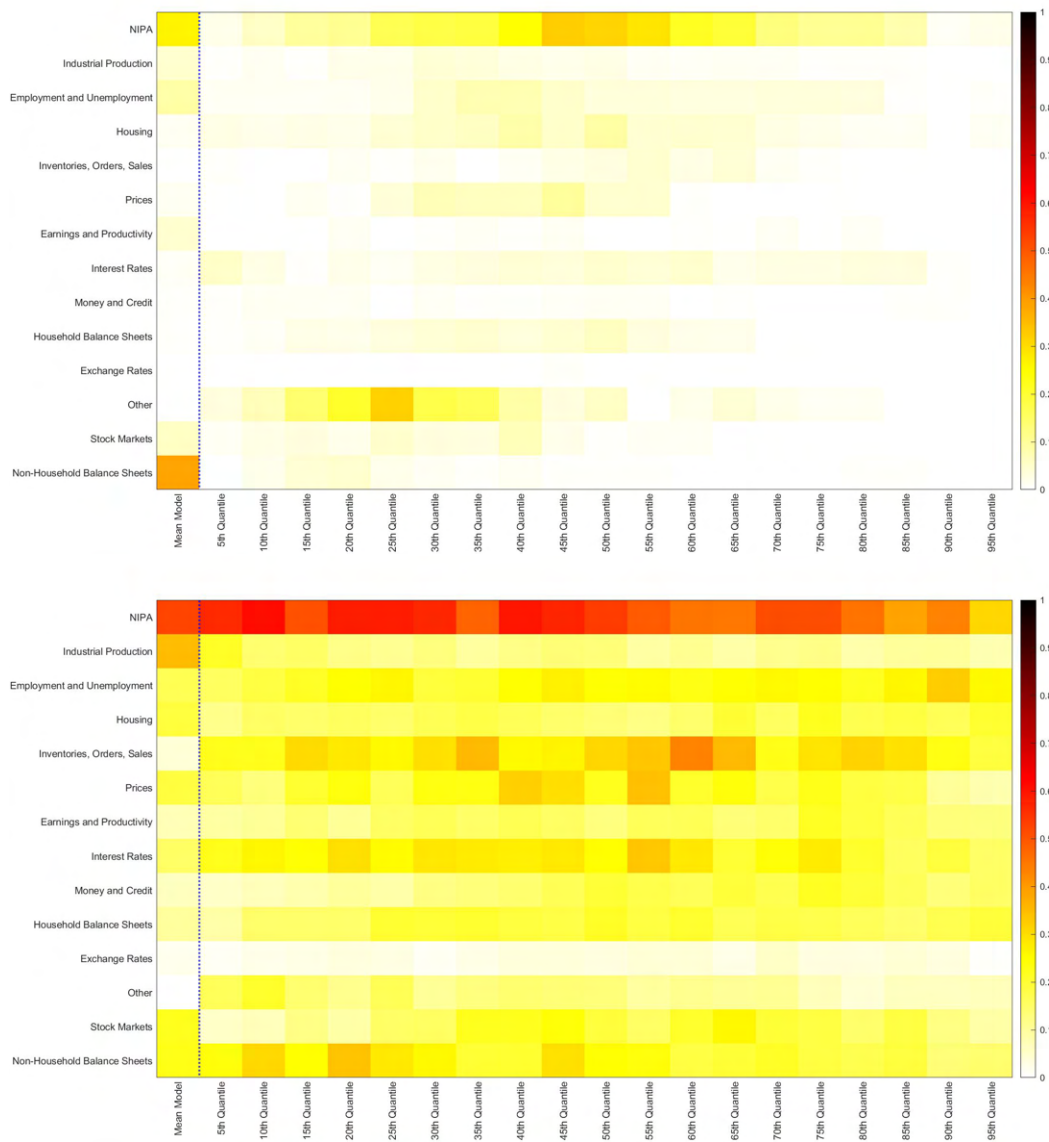


Figure 5.7: Maximum (per data group) of average inclusion probability for each quantile of the $LBQR_{BIC}$ (top) and the $SSVSBQR$ (bottom) for $h=1$

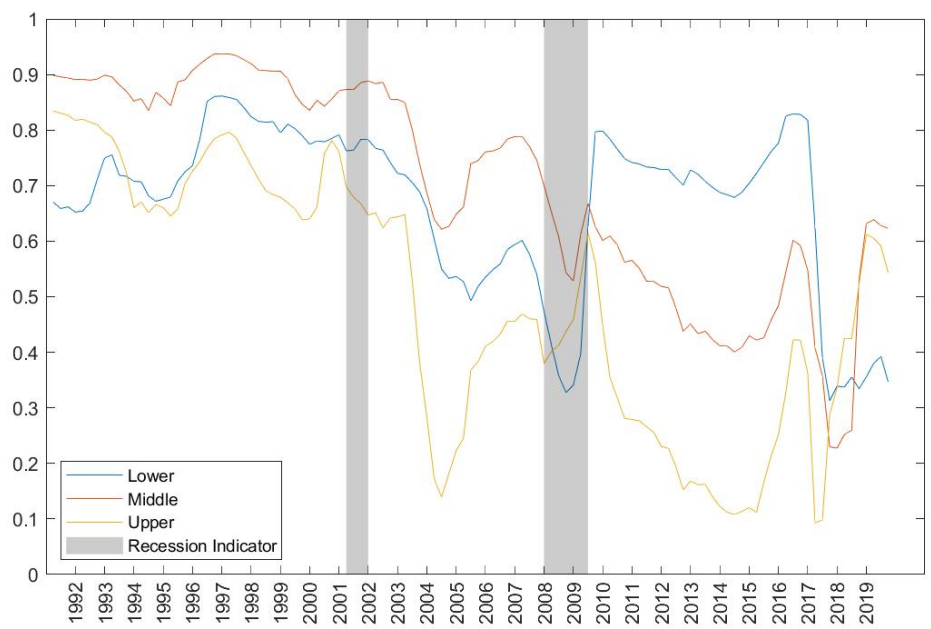
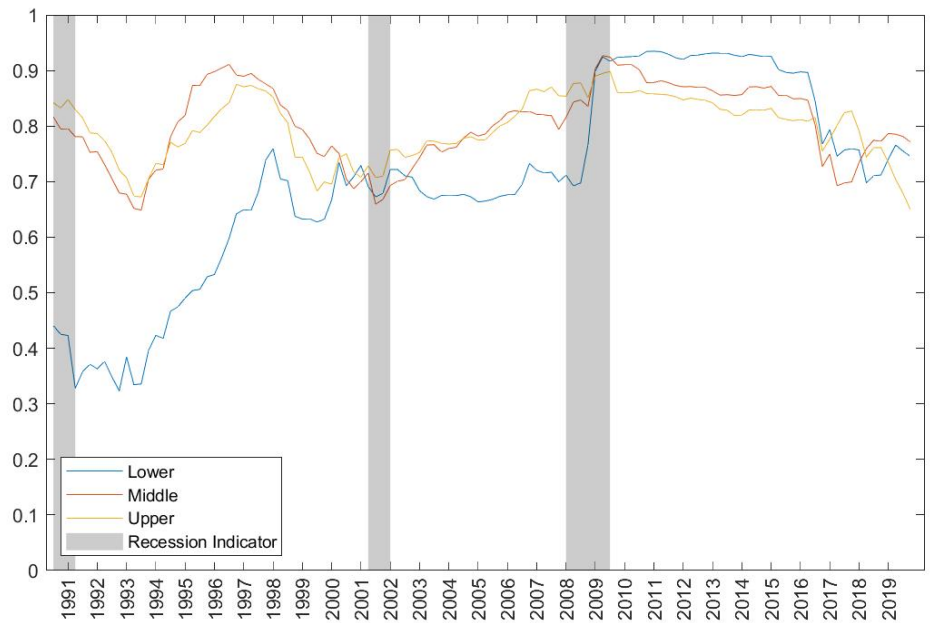


Figure 5.8: Correlation coefficients computed on a rolling 30 quarter window

than under the horseshoe prior, while the $SSVSBQR$ tends to select larger models (lower panel of figure 5.7). This is corroborated by table (5.6) which shows average model sizes of selected estimators. Lastly, the $SSVSBQR$ prior also induces far greater variable inclusion uncertainty, as the heatmap shows denser yellow regions compared to the $HSBQR_{BIC}$. Nevertheless, variables with the highest inclusion probabilities tend to overlap with those selected by the $HSBQR_{BIC}$.

The fact that the $SSVSBQR$ prior increases model uncertainty without outperforming the $HSBQR$ and $HSBQR_{BIC}$ might indicate that the discretisation of the model-space along with the less flexible slab distribution employed by the $SSVSBQR$ prior induces higher uncertainty over variable inclusion, especially with correlated data. This, in fact, corroborates recent work by Fava and Lopes (2021) who find that when allowing for slab distributions with fatter tails than the normal, posterior variable inclusion uncertainty can be greatly reduced. Using global-local priors, and bringing our sparsification approach to the quantile setting, we give new impetus to the debate on the “illusion of sparsity”.

Variable Inclusion Across Time

While variable inclusion averaged across time revealed that there is support for quantile varying sparsity in the high dimensional GaR, figures (5.9) and (5.10) clearly show that variable selection has substantial time variation as well. Taking the $HSBQR_{BIC}$ as the benchmark, we can see in figure (5.9) that the time variation in variable selection is in line with what one would expect given the economic history of the US. There are two periods that stand out at one-quarter ahead forecast: the financial crisis of 2008 and the mid to late 80’s. Around the financial crisis the left tail increasingly includes employment and financial data such as interest rates and non-household balance sheet information. During the mid-to late 80’s, a period which is often referred to as the Volcker period in the macro literature²¹, price variables have high inclusion probability. The middle and right tail share many common variable inclusion patterns over time, where employment and housing data impact the selected models during the 2000’s and the financial crisis.

²¹ When Paul Volcker became chairman of the Federal Reserve Board in 1979, the annual average inflation rate in the United States was 9%. The Federal Reserves was able to bring inflation down to 4% by the end of 1983. However, this disinflation came at a cost as during this period, the US experienced two recessions that are attributed to disinflationary monetary policy (Goodfriend and King, 2005).

	Left	Mid	Right	Left	Mid	Right
		h=1			h=4	
$SSVSBQR$	13.023	19.195	14.668	9.063	13.024	8.505
$HSBQR_{BIC}$	9.003	9.640	8.202	7.474	7.314	7.148
$LBQR_{BIC}$	1.541	3.055	1.107	1.439	2.748	1.119

Table 5.6: Average Model sizes of the different models

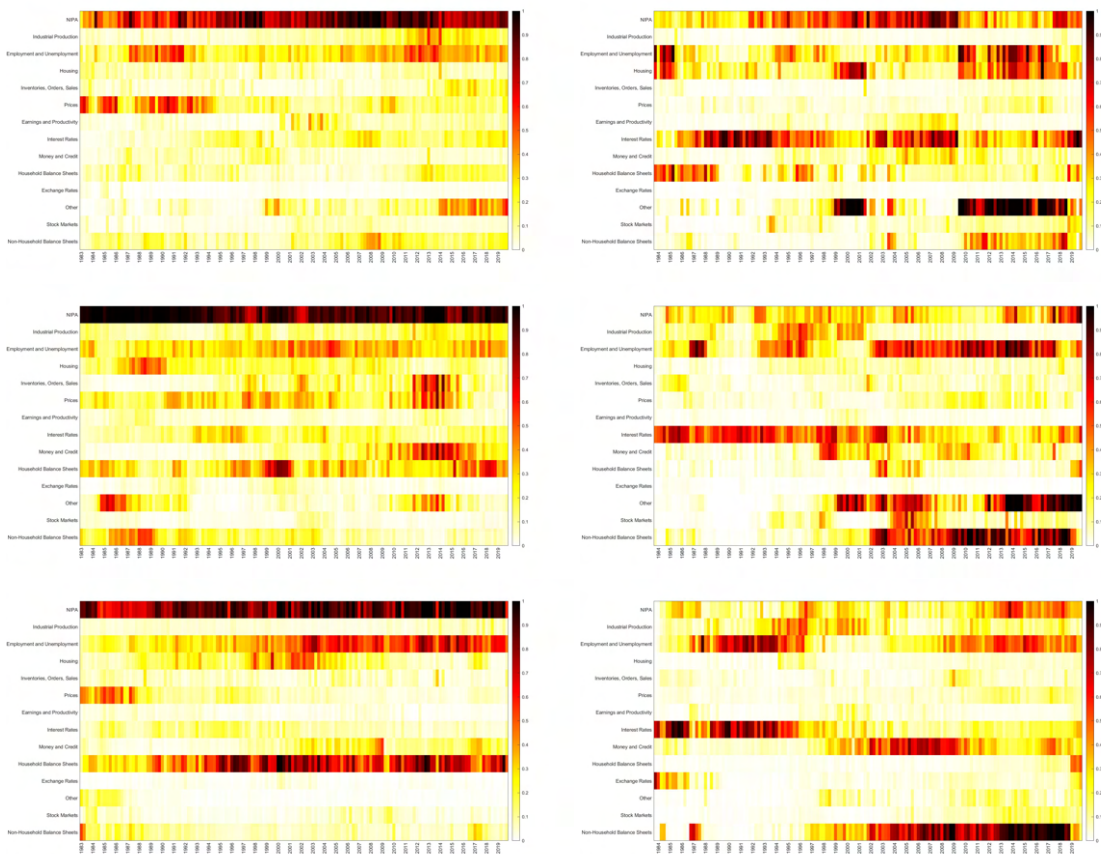


Figure 5.9: Maximum (per data group) of average inclusion probability each year for the $HSBQR_{BIC}$ for its 3 left quantiles (top), 3 central quantiles (middle), and 3 right quantiles (bottom) for $h=1$ (left) and $h=4$ (right)

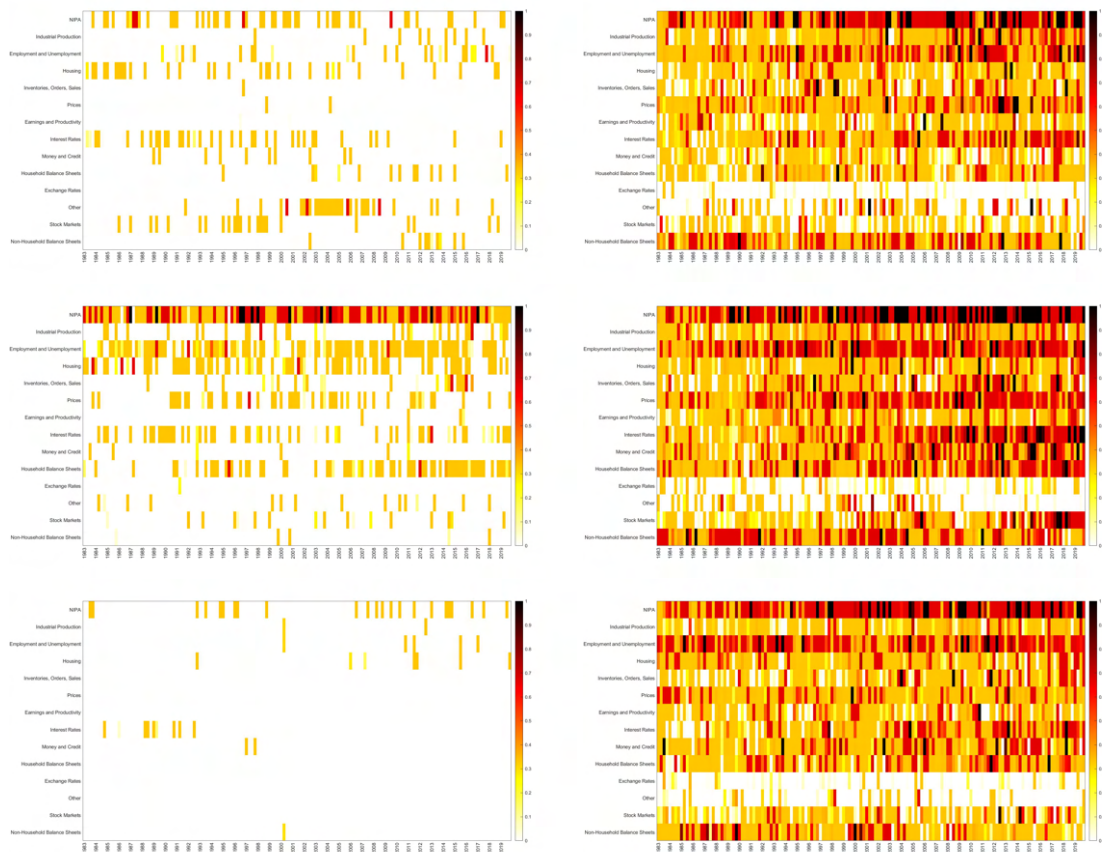


Figure 5.10: Maximum (per data group) of average inclusion probability each year for the $LBQR_{BIC}$ (left) and $SSVSBQR$ (right) for their 3 left quantiles (top), 3 central quantiles (middle), and 3 right quantiles (bottom) for $h=1$

Comparing to the $LBQR$ and $SSVSBQR$ in figure (5.10), the contrasts in variable selection over time are even starker. The $LBQR_{BIC}$ produces very sparse and the $SSVSBQR$ models produce denser models, and the inclusion probabilities for both models are very erratic. Seldom is a variable included with high probability for either priors for more than a quarter which can be seen from the very dotted heatmaps. The $HSBQR$ models on the other hand produce far smoother inclusion patterns which indicates that variables with high inclusion probability impact forecasts for longer periods. While in figure (5.10) we present $LBQR_{BIC}$, the same is true for $LBQR_{SAVS}$.

The general finding that there is time variation in the importance of variables for aggregate macroeconomic time-series has already been extensively studied for conditional mean models such as in Koop and Korobilis (2012), Billio et al. (2013) and McAlinn and West (2019). Our findings motivate that similar methods be also considered for the Bayesian quantile regression moving forward. It is however not surprising that quantile regression models are more sensitive to varying variable importance, even with static parameters. Namely, as the time-series accumulates, crisis periods identify new information for the tails which the tick-loss function will heavily tilt the fitted quantiles toward. This explains the large increase in the importance of variable groups which help predict the 2000 and financial crisis, especially for the left tail and the median.

5.8 Conclusion

This chapter has, in the spirit of Hahn and Carvalho (2015) and Ray and Bhattacharya (2018), proposed to conduct variable selection for continuous shrinkage priors within the Bayesian Quantile Regression by decoupling shrinkage from sparsity as derived from a Bayesian decision theoretic perspective. The resultant easy to implement SAVS procedure for the BQR selects variables on a quantile specific basis where the degree of sparsity is estimated from the data via the quantile BIC in order to allow for heterogeneous levels of penalisation across the conditional distribution.

Applying the proposed method to both simulated experiments and an empirical application, we have found that the qBIC augmented version, the BQR_{BIC} , retains or even improves the fit of the un-sparsified posterior, while giving easy to interpret variable selection results.

Simulated experiments have tested a variety of sparsity patterns and error distributions, both for quantile constant and quantile varying coefficient vectors and have found that sparsification often lowers bias, or at worst holds bias stable, while improving variable selection compared to the traditional Bayesian variable selection priors such as the SSVS. The proposed selection procedure via the qBIC outperforms the BQR_{SAVS} especially in sparse DGPs and more generally in the tails where the

extra flexibility of determining the penalisation from the data allows to capture the quantile profile.

The high dimensional growth-at-risk application to the US verifies the findings of the Monte Carlo exercise, namely that the BQR_{BIC} retains density forecast performance and even improves downside risk estimation. Compared to the benchmark ‘vulnerable growth’ model of Adrian et al. (2019), the big data models provide significant gains in density forecasts. In particular, the proposed sparsification techniques applied to the $HSBQR$ provide well calibrated densities with much improved left tail performance. Variable selection results of the proposed methods revealed that there is substantial quantile varying sparsity for US macroeconomic data. As intuition would suggest variable groups which predict crises periods well, such as interest rate and housing data appear with higher inclusion probability in the left tail while right and middle quantiles tend to be affected by many overlapping variable groups.

Contrasting these results to variable selection of more commonly used conditional mean models, we find that mean models tend to select variable groups which have high inclusion probability not only in the median, but also the tails. This highlights the importance of quantile models as it allows us to decompose which variables impact different parts of the distribution. The general finding that there is time variation in the importance of variables for aggregate macroeconomic time-series seems to hold true for quantile models as well. We feel that it is a worthy endeavour to further research methods that allow for quantile as well as time variation for variable selection, to gain a better understanding of which variables are important for the different quantiles across time.

Chapter 6

Global Local Priors and The Illusion of Sparsity

6.1 Introduction

Recent decades within social and economic research have experienced unprecedented increases in data availability which resulted in an equal amount of demand in the development of statistical and econometric methods that can deal with vast amounts of information; often dubbed machine learning (ML) or Big Data methods. Their necessity is prompted by the fact that OLS methods are not adequate or even possible to estimate in situations where there are more covariates than observations (Tibshirani, 1996). To make estimation feasible, ML and Big Data methods perform regularisation which forces the effects of some variables to zero; thereby freeing up degrees of freedom. While a great variety of regularisation methods have been proposed, both linear and non-linear (see for an overview Hastie et al. (2009)), for predictive purposes, the economic literature still places great importance on extensions to linear regression methods due to their often found competitive performance compared to more complicated non-linear methods (e.g. neural nets, support vector machines, random forests) (Coulombe et al., 2020), and more importantly their interpretability, in that the partial effects from a regression model can more easily be communicated to policy makers.

Following Ng (2013), Chernozhukov et al. (2017) and Giannone et al. (2021), linear prediction methods that deal with high dimensions can be categorised into either sparse or dense modeling approaches. Sparse models such as lasso (Tibshirani, 1996), adaptive lasso (Zou, 2006), SCAD (Fan and Li, 2001) embed the assumption of sparsity through a penalisation term in the minimisation process or the prior which forces most of the entries in the regression vector to zero, while dense models such as ridge regression (Hoerl and Kennard, 1970) and PCA methods (Stock and Watson, 2002) assume that most of the variables included have small non-zero effects. This is often modeled via a rotation of the original variable space into their

first r principal components.

The specific modeling choice often adheres to literature specific conventions. In the macroeconomic literature, dense approaches that summarise common movement in large sets of macro indicators such as factor and dynamic factor models (DFM) are considered workhorse models due to a deep structural relationship between business cycle dynamics and the co-movement of aggregate macroeconomic data (Stock and Watson, 2002). DFMs are also popular for nowcasting where predictions of such models are often used as inputs into more structural models such as DSGE models (Giannone et al., 2016). In the microeconomic literature on the other hand, many studies use the lasso or variants of the lasso (Belloni et al., 2012) in order to identify policy relevant intervention effects. This in particular highlights that the model choice and the assumption about the degree of sparsity not only has implications for the quality of predictions, but also on how we interpret the data.

An open question in the literature is whether the data can be informative about which modeling technique is preferred. As such, recent contributions such as Giannone et al. (2021) and Fava and Lopes (2021) investigate through Bayesian spike and slab priors whether the data are informative about the degree of sparsity. The spike and slab prior assumes a discrete mixture of normals prior on the regression coefficients which is augmented by the above authors through $\mathcal{B}(\bullet, \bullet)$ priors on the probability of inclusion parameter, q , and a separate prior on a shrinkage scale, γ , so as to stay agnostic a-priori about the degree of sparsity and the amount of shrinkage. The advantage of the Bayesian approach over frequentist lasso style estimators, is that one obtains statistically coherent uncertainty measures on q and γ . Fava and Lopes (2021) extend this model further by allowing for a fat-tailed t-distributed slab distribution so as to relax the amount of shrinkage on potentially large coefficients. By applying their prior to a host of economic data sets, they find that there is substantial uncertainty in the degree of sparsity for most data sets which is hidden in methods that take a stand on the degree of sparsity a-priori.

Drawbacks of these studies are firstly that spike and slab priors rely on strict parametric assumptions about the shrinkage distribution (normal in the case of Giannone et al. (2021) and t-distribution in the case of Fava and Lopes (2021)). Secondly, due to combinatorial bottlenecks, discrete mixture priors such as the spike and slab rely on approximations to the model space which can cause mixing problems, particularly in high dimensional (Barbieri et al., 2021) settings and with correlated data (Ishwaran et al., 2005; Malsiner-Walli and Wagner, 2018; Piironen et al., 2020). It is additionally well known that the spike and slab prior is sensitive to the parameterisation of the slab distribution and selection of the hyper-parameters (Barbieri and Berger, 2004; O’Hara and Sillanpää, 2009; Fava and Lopes, 2021). This can result in false certainty about the posterior degree of sparsity. It may also be argued

that measures of the degree of sparsity based on marginal likelihoods, as is the case for spike-and-slab priors, are inappropriate when the policymaker maximises the joint goal of out-of-sample predictive performance (Piironen et al., 2020).

Instead, the broader Bayesian literature has recently popularised continuous shrinkage priors such as the horseshoe Carvalho et al. (2010), regularised horseshoe Piironen et al. (2017), R2D2 Zhang et al. (2020), Dirichlet-Laplace Bhattacharya et al. (2015) and Normal-Gamma Griffin et al. (2010) prior which offer computational advantages due to their continuous formulation as well favourable asymptotic properties (Van Der Pas et al., 2014; Zhang et al., 2020). Cross et al. (2020) and Kohns and Bhattacharjee (2022) show that global-local priors offer competitive forecasting performance in macroeconomic applications. However, using continuous global-local priors for sparsity detection in economic applications has previously been hampered by the fact that also the posterior of the regression parameters remain non-sparse (Hahn and Carvalho, 2015). In this chapter, we investigate whether global-local priors are able to uncover the degree of sparsity in data informed fashion and further compare their forecasting performance in large scale Monte Carlo simulations as well as a host of economic data sets.

In order to infer on the degree of sparsity of global-local priors, we make use of two methods: (1) an effective model size measure based on a shrinkage coefficient distribution representation of the posteriors (2) search algorithms that sparsify the posterior of the GL priors. The effective model size measure has been introduced by Piironen et al. (2017) in the context of the regularised horseshoe prior, which measures the count of active regressors in the posterior and is used by the authors to construct an informative hyperparameter on the global shrinkage distribution. To the best of our knowledge we are the first to apply this measure to other global-local priors and to use it to investigate the degree of sparsity. Sparsification of the regression posterior on the other hand is based on the idea that regression coefficients can be thresholded to zero, when the impact on the response is small. Recently proposed methods for the latter purpose are the SAVS algorithm of Ray and Bhattacharya (2018) which uses lasso style heuristics under squared error loss and the predictive projective inference approach by Piironen et al. (2020) which gives fully Bayesian sparsification under Kullback-Leibler loss and uses Bayesian leave-one-out cross-validation for subset selection based on out-of-sample performance. While the SAVS algorithm is becoming increasingly popular (e.g. Huber et al. (2019); Kohns and Bhattacharjee (2022); Chakraborty et al. (2020); Hauzenberger et al. (2020)), the forecasting and sparsification properties of predictive projective inference are largely unexplored for economic data.

Motivated by the findings of our analysis, we propose the agnostic regularised

horseshoe (A-RHS) prior, which adds a prior over the expected ratio of non-zero coefficients to the total covariate dimension that allows the data to speak about the preferred degree of sparsity. With this, we provide an alternative to the spike-and-slab variable selection approach with uninformative priors on variable inclusion.

The theoretical analysis shows that most global-local priors are in fact informative about the implied degree of sparsity. We show via simulations and real data applications that less flexible global-local priors tend to carry over their sparsity preference also to their posterior distributions. By controlling our prior beliefs about expected model sizes via an ‘agnostic’ Beta prior extension on the expected ratio of non-zero to total coefficients to the regularised horseshoe prior, we are able to remedy this behaviour. Our simulations show that the the agnostic RHS as well as regularised horseshoe prior with correct guesses on sparsity provide competitive inference on sparsity compared to spike-and-slab priors while being able to offer similar or slightly better forecasting performance. The good forecasting performance is confirmed in a host of economic data sets and contributes to the findings Fava and Lopes (2021) that the spike-and-slab prior with a normal slab distribution gives far larger uncertainty about the degree of sparsity than our A-RHS and competing global-local priors.

Sparsification approaches confirm our analysis based on implicit model size distributions and highlight that the quality of shrinkage will heavily influence the ability of either sparsification approach to detect signal from noise. The empirical evidence suggests that the extra flexibility of projective inference to use out-of-sample predictive performance for variable selection is often preferable over the SAVS algorithm which tends to over-sparsify, particularly in dense DGPs.

In the following, we firstly review the theory behind effective model sizes as well as SAVS and projective predictive inference. Then, we conduct simulations and present results from the empirical applications. Finally, we conclude.

6.2 Theory of Model Sizes

6.2.1 A General Framework for Implicit Model Sizes

Global-local priors are continuous shrinkage priors which share the trait that their marginal prior distribution on the regression coefficients provide large mass on zero (typically far more than the normal distribution) and fat tails which allow for signals with large coefficients to escape shrinkage (Polson and Scott, 2010). The general

class of global-local priors take the following form:

$$\begin{aligned}\beta_j | \varphi_j^2, \vartheta^2, \sigma^2 &\sim N(0, \varphi_j^2 \vartheta^2 \sigma^2), j \in (1, \dots, K) \\ \varphi_j &\sim \pi(\varphi_j), j \in (1, \dots, K) \\ \vartheta &\sim \pi(\vartheta)\end{aligned}\tag{6.1}$$

The idea of this family of priors is that the global scale ϑ controls the overall shrinkage applied to the regression, while the local scale φ_j allows for the local possibility of regressors to escape shrinkage if they have large effects on the response. Unlike discrete mixture spike and slab priors such as (Giannone et al., 2021; Mitchell and Beauchamp, 1988), (6.1) does not provide a direct measure of sparsity, as the prior does not provide positive probability on 0. Instead, we leverage an implied posterior model size formulation of Piironen et al. (2017) which defines the implied model distribution directly over the conditional posterior of the regression coefficients.

Consider a linear Gaussian regression with T observations:

$$y_t = x_t' \beta + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma^2), \quad \forall t = 1, \dots, T\tag{6.2}$$

with fixed covariate matrix X , of dimension $(T \times K)$ and a K dimensional static regression vector, β . Assuming a scale mixture of normal global-local prior for (β, σ) , the conditional posterior β can be expressed as¹:

$$\begin{aligned}p(\beta | \Lambda, \vartheta, \sigma^2, y) &= N(\bar{\beta}, \Sigma) \\ \bar{\beta} &= \vartheta^2 \Lambda (\vartheta^2 \Lambda + \sigma^2 (X'X)^{-1})^{-1} \hat{\beta} \\ \Sigma &= (\vartheta^2 \Lambda^{-1} + \frac{1}{\sigma^2} X'X)^{-1},\end{aligned}\tag{6.3}$$

where $\Lambda = \text{diag}(\varphi_1^2, \dots, \varphi_K^2)$ and $\hat{\beta}$ is the maximum likelihood solution to (6.2), assuming that X is full column rank. Assuming for simplicity that the covariates are zero mean and uncorrelated with columnwise standard deviation, s_j , (6.3) can be shown to reduce to:

$$\bar{\beta}_j = (1 - \kappa_j) \hat{\beta}_j\tag{6.4}$$

where

$$\kappa_j = \frac{1}{1 + T \sigma^{-2} \vartheta^2 s_j^2 \varphi_j^2}.\tag{6.5}$$

κ_j can be regarded as a shrinkage coefficient as it's 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 $\varphi_j \vartheta \rightarrow \infty$, then $\kappa_j \rightarrow 0$ and when $\varphi_j \vartheta \rightarrow 0$, then $\kappa_j \rightarrow 1$. The distribution of the shrinkage coefficient $p(\kappa_j | \Lambda, \vartheta, \sigma^2, y)$ is implicitly defined through the priors for the hyperparameters (Λ, ϑ) . Piironen et al. (2017)

¹ Note, that we are assuming dependence in the conditional prior of β and σ . Similar derivations can be conducted under the assumption of independence.

propose a joint measure of the effective model size (m_{eff}) which can be derived using the fact that the posterior of κ_j is independent given the hyperparameters:

$$m_{eff} = \sum_{j=1}^K (1 - \kappa_j). \quad (6.6)$$

Hence, when the shrinkage coefficients are distributed with most mass on 0 or 1, as is the case for most global-local priors, then the distribution of this quantity over all MCMC draws gives an indication as to how many coefficients remain unshrunk and therefore the implied level of sparsity. While Piironen et al. (2017) use this formulation to derive a scaling factor for the regularised horseshoe prior, further discussed below, we will use this formulation to quantify the degree of sparsity global-local priors prefer.

A word of caution is in order at this point. This chapter abstracts from purely frequentist properties of global local priors considered in papers such as Van Der Pas et al. (2014), Bhattacharya et al. (2015) or Zhang et al. (2020) who derive the asymptotic behaviour and contraction rates of the posteriors under various assumptions about the growth of non-zero coefficients in the true parameter vector β_0 . This chapter, in contrast, is concerned with the finite sample performance of the priors in detecting signals from noise for sparsity detection as well as prediction problems and how these compare to established discrete mixture of normals priors. We now investigate the shrinkage coefficient distribution for the GL priors under consideration and provide prior predictive distribution according to (6.6) to establish how the properties of the global-local priors influence the effective model size distribution a-priori.

6.2.2 Deriving Implicit Model Size Distributions

For this section and the rest of the chapter, we restrict our analysis to a selection of global-local priors which have garnered much attention in the statistics as well as economic literature, namely the horseshoe prior (HS) of Carvalho et al. (2010), regularised horseshoe prior (RHS) Piironen et al. (2017), Dirichlet-Laplace prior (DL) (Bhattacharya et al., 2015), Normal-Gamma prior (NG) and R2D2 prior (R2D2) of Zhang et al. (2020). The popularity of these and related priors² stem from their favourable theoretical properties and good empirical performance compared to other global-local prior classes³.

The widely popular horseshoe prior by Carvalho et al. (2010) employs two half Cauchy

² See for example Bhadra et al. (2019) and Bhadra et al. (2017) for an overview of competing priors.

³ See Zhang et al. (2020) for an extensive discussion.

distributions on the positive support for the global and local scales:

$$\begin{aligned}\beta_j|\varphi_j^2, \vartheta^2 &\sim N(0, \varphi_j^2\vartheta^2) \\ \varphi_j &\sim C_+(0, 1), \quad j \in (1, \dots, K) \\ \vartheta &\sim C_+(0, 1),\end{aligned}\tag{6.7}$$

where $C_+(0, 1)$ denotes the density $\pi(y|\sigma) = \frac{2}{\pi\sigma(1+\frac{y}{\sigma})^2}$.

The regularised horseshoe prior of Piironen et al. (2017) extends on the horseshoe by addressing two problems related to the Cauchy tails and the scale of the global shrinkage distribution. Since the Cauchy distribution has undefined moments which can lead to vanishing posterior means (Ghosh et al., 2018), Piironen et al. (2017) add a hierarchical scale, c , to the local shrinkage distribution of φ_j which effectively acts to soft truncate the implied distribution of κ_j from zero to $\tilde{\kappa}_j = 1/(1 + T\sigma^{-2}s_j^2c^2)$:

$$\begin{aligned}\beta_j|\varphi_j^2, \vartheta^2, c &\sim N(0, \vartheta^2, \tilde{\varphi}_j^2), \quad \tilde{\varphi}_j^2 = \frac{c^2\varphi_j}{\frac{\sigma^2}{Ts_j^2} + c^2 + \vartheta^2\varphi_j^2} \\ \varphi_j &\sim C_+(0, 1) \quad j \in (1, \dots, K) \\ \vartheta &\sim C_+(0, \vartheta_0) \\ c &\sim \mathcal{G}^{-1}(\eta_1, \eta_2), \quad \eta_1 = \nu_c/2, \quad \eta_2 = \nu_c s_c^2/2,\end{aligned}\tag{6.8}$$

where $i\mathcal{G}$ refers to the Inverse-Gamma distribution with density $\pi(y) = \frac{\beta^\alpha}{\Gamma(\alpha)}y^{-\alpha-1}\exp(-\frac{\beta}{y})$.

The addition of c retains the form of the shrinkage coefficient distribution in that it is unbounded with singularity at $\kappa_j = 1$, however regularises the amount to which β_j can be estimated freely. Piironen et al. (2017) further extend on the scale ϑ_0 , which they show has a substantial impact on the effective model size distribution (6.6). In particular, through the change of variables theorem, it can be shown that the horseshoe prior (6.7) implies the following distribution on κ_j :

$$p(\kappa_j|\vartheta, \sigma) = \frac{1}{\pi} \frac{a_j}{(a_j^2 - 1)\kappa_j + 1} \frac{1}{\sqrt{\kappa_j}\sqrt{1 - \kappa_j}},\tag{6.9}$$

where $a_j = \vartheta\sigma^{-1}\sqrt{T}s_j$. When $a_j = 1$, (6.9) reduces to a Beta(0.5,0.5) distribution which results in the name giving horseshoe shape in figure (6.2). Piironen et al. (2017) derive analytical expressions for $E(\kappa_j|\vartheta, \sigma)$ and $Var(\kappa_j|\vartheta, \sigma)$, which, combined with equation (6.6), results in the expected model size and its variance:

$$E(m_{eff}|\vartheta, \sigma) = \sum_{j=1}^K \frac{a_j}{1 + a_j}\tag{6.10}$$

$$Var(m_{eff}|\vartheta, \sigma) = \frac{a_j}{2(1 + a_j)^2}\tag{6.11}$$

Plugging back in $a_j = \vartheta\sigma^{-1}\sqrt{T}s_j$, and assuming a prior model size guess, κ_0 , the effect of the global local scale, ϑ_0 , can be intuitively understood as the fraction of active predictors scaled by the level of noise, σ , and the number of observations:

$$\vartheta_0 = \frac{\kappa_0}{K - \kappa_0} \frac{\sigma}{\sqrt{T}}. \quad (6.12)$$

For all subsequent analysis, we follow the recommendations by the Stan community Carpenter et al. (2017) and set $\nu_c = 4, s_c = 1$.

Since this chapter is concerned with letting the data speak about the effective model size, we extend on the regularised horseshoe prior by adding a $\mathcal{B}(\alpha_2, \beta_2)$ distribution to the fraction of active coefficients, κ_0 , to total covariate dimension K , which we define as $\rho = \frac{\kappa_0}{K}$. By slightly reformulating (6.12) in terms of ρ , the prior has the following hierarchy:

$$\begin{aligned} \beta_j | \varphi_j^2, \vartheta^2, c &\sim N(0, \vartheta^2, \tilde{\varphi}_j^2), \quad \tilde{\varphi}_j^2 = \frac{c^2 \varphi_j}{\frac{\sigma^2}{Ts_j^2} + c^2 + \vartheta^2 \varphi_j^2} \\ \varphi_j &\sim C_+(0, 1) \\ \vartheta &\sim C_+(0, \vartheta_0), \quad \vartheta_0 = \frac{\rho}{1 - \rho} \frac{\sigma}{\sqrt{T}} \\ c &\sim \mathcal{G}^{-1}(\eta_1, \eta_2), \quad \eta_1 = \nu_c/2, \quad \eta_2 = \nu_c s_c^2/2 \\ \rho &\sim \mathcal{B}(\alpha_2, \beta_2) \end{aligned} \quad (6.13)$$

Throughout simulations and applications, we set $\alpha_2 = \beta_2 = 1$ to allow any fraction of active regressors to be a-priori equally likely.

Theorem 1. *Let $\rho \sim \mathcal{B}(\alpha_2, \beta_2)$, then $\vartheta_0 \propto \beta'(\alpha_2, \beta_2)$, where $\beta'(\alpha_2, \beta_2)$ refers to the beta prime distribution, whose probability density function is defined as $\frac{x^{\alpha_2-1}(1+x)^{-\alpha_2-\beta_2}}{B(\alpha_2, \beta_2)}$, and $B(\alpha_2, \beta_2)$ denotes the beta function.*

The proof for the theorem follows immediately from application of the change of variable theorem which implies that if $X \sim \mathcal{B}(a, b)$, then $\frac{X}{1-X} \sim \beta'(a, b)$ (Johnson et al., 1995). We name this prior the agnostic horseshoe prior (A-RHS). The rest of the hyperparameters for the A-RHS are set similarly to the RHS.

The Dirichlet-Laplace prior of Bhattacharya et al. (2015), belongs to a class of priors that instead of having a single global shrinkage prior, allow the global scale to vary with each coefficient, with the restriction that the global scale process follows a simplex structure. This is enforced through a Dirichlet prior which is shown by (Bhattacharya et al., 2015) to imply shrinkage on the cumulative effects of the regressors. Next to achieving optimal near-minimax rates of posterior contraction (Bhattacharya et al., 2015), the simplex structure induces dependence in shrinkage

across β_j . In particular, the DL prior assume:

$$\beta_j | \sigma, \psi_j, \phi, \omega \sim N\left(0, \psi_j \phi_j^2 \omega^2 \frac{\sigma^2}{2}\right) \quad (6.14)$$

$$\psi_j \sim \exp(1/2), \phi \sim \text{Dir}(a_\phi, \dots, a_\phi), \omega \sim \mathcal{G}(Ka_\phi, 1/2)^4,$$

where $\exp()$ and $\text{Dir}()$ refer to the exponential and Dirichlet distribution respectively. We set the hyperparameter $a_\phi = 0.5$ for all subsequent analysis which is recommended by Bhattacharya et al. (2015) to be a good default choice.

The recently proposed R2D2 prior by Zhang et al. (2020) belongs to the class of priors motivated by a prior on the coefficients of determination, R^2 , rather than on the regression coefficients directly. The authors show that a $\mathcal{B}(\bullet, \bullet)$ prior on R^2 , implies a regularising prior on the squared formulation $\beta' X' X \beta$, which, paired with a local scale shrinkage prior induces the following prior hierarchy after marginalising out the design:

$$\beta_j | \sigma^2, \psi_j, \phi, \omega \sim N\left(0, \psi_j \phi_j \omega \frac{\sigma}{2}\right) \quad (6.15)$$

$$\psi_j \sim \exp(1/2), \phi \sim \text{Dir}(a_\phi, \dots, a_\phi), \omega | \xi \sim \mathcal{G}(Ka_\phi, \xi), \xi \sim \mathcal{G}(b_\phi, 1).$$

Similar to the DL prior, global shrinkage is adaptive across j and adds an additional hierarchy to the global scale, ω , whose rate follows a gamma distribution. This prior, too, converges at the optimal near-minimax rate. As recommended by Zhang et al. (2020), we set $b_\phi = 0.5$ and $a_\phi = 1/(K^{b/2} T^{b/2} \log(T))$ which by theorem 6 in Zhang et al. (2020) is required for posterior consistency.

Finally, the popular Normal-Gamma prior of Griffin et al. (2010) assumes the following prior setup:

$$\beta_j | \varphi_j, \vartheta \sim N\left(0, \frac{2}{\vartheta^2} \varphi_j\right) \quad (6.16)$$

$$\varphi_j \sim \mathcal{G}(\theta, \theta), \theta \sim \exp(1), \vartheta \sim \mathcal{G}(c_0, c_1).$$

This prior generalises the lasso prior of Park and Casella (2008) which is recovered by fixing $\theta = 1^5$.

6.2.3 Shrinkage Coefficient Distributions

With priors (6.7)-(6.16) at hand, we provide in table (6.1) the implied shrinkage coefficient distributions. To the best of our knowledge these have not yet been derived

⁴ Notice that we are switching notation slightly to what is commonly used for the exposition of those priors (Zhang et al., 2020)

⁵ All priors discussed above have marginal prior distributions for β_j which are unbounded with a singularity at zero and different concentration rates as well as tail decay rates. Zhang et al. (2020) show that the R2D2 has the fastest concentration rate at 0 (of polynomial order in comparison to the logarithmic order of the horseshoe prior variants) as well as the heaviest tail with the slowest tail decay (again of polynomial order compared to the horseshoe's β^2).

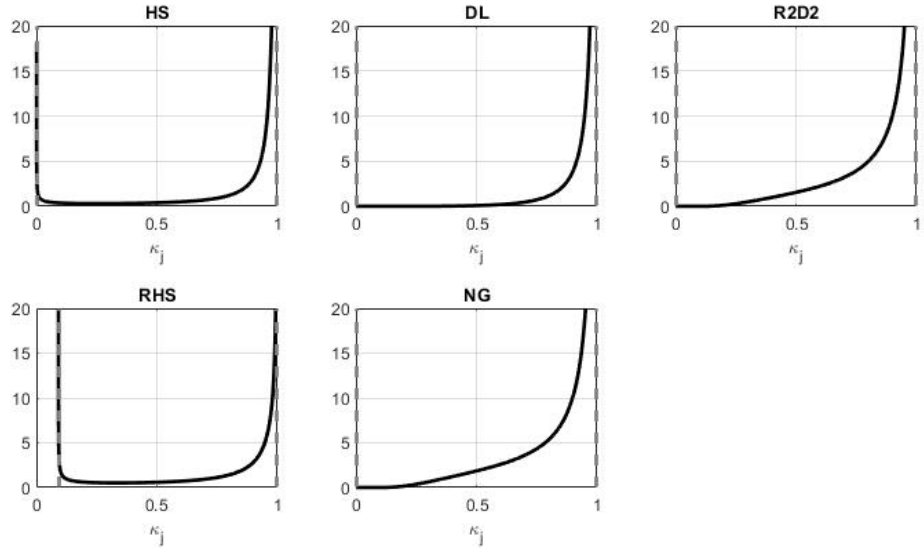


Figure 6.1: Shrinkage Distributions for the HS, DL, R2D2, RHS and NG prior respectively. For the RHS prior, we set $\kappa_0 = 1$. a_j is set to 0.1 which corresponds to strong global shrinkage, thereby favouring sparse solutions. a_j is computed for each prior according to table (6.1).

for the DL and R2D2 prior. Since $p(\kappa_j|\vartheta, \sigma)$ is derived conditional on the global shrinkage parameters and the error variance, we therefore present the shrinkage distributions for instances for which the global scale is set such that a_j solves for 0.1, 1 and 10. This represents the global shrinkage term indicating sparsity, moderate sparsity and moderate density respectively.

For $a_j = 0.1$ in figure (6.1), which represents strong global shrinkage, one can see that all global-local priors considered, have most mass $\kappa_j = 1$ which favours complete shrinkage. Noticeably, the horseshoe based priors provide not only singularity at 1, but also retain some mass around $\kappa_j = 0$, albeit with very quick convergence to the boundary. Hence, while most predictors will be shrunk to zero, the HS can still accommodate large coefficient values given strong signals. The other priors, on the other hand still tend to shrink large signals. By decreasing the degree of sparsity in figures (6.2) and (6.3), the DL, R2D2 and NG prior converge to shrinkage distributions which follow a Cauchy like U-shape, therefore accommodating more mass on no shrinkage, with however considerable mass on $\kappa_j = 1$, thereby still exerting some shrinkage. The horseshoe priors reverse the shape compared to $a_j = 0.1$ with most mass on no shrinkage, thereby favouring most regression coefficients be completely unshrunk. In line with these shrinkage coefficient properties, Zhang et al. (2020) find that the horseshoe has the lowest bias for large coefficients, while the DL and R2D2 perform better for small coefficients.

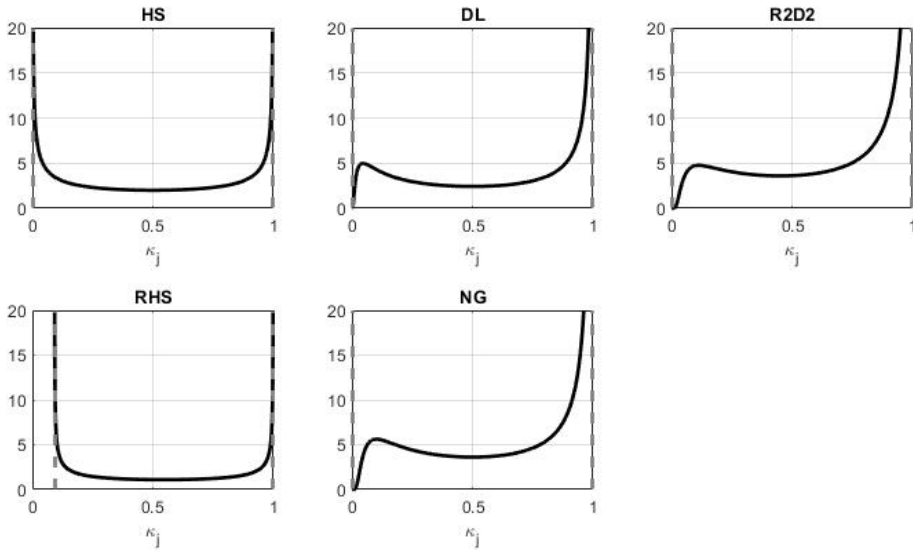


Figure 6.2: Shrinkage Distributions for the HS, DL, R2D2, RHS and NG prior respectively. For the RHS prior, we set $\kappa_0 = 1$. a_j is set to 1 which corresponds to medium strong global shrinkage, thereby enabling both sparse and dense solutions. a_j is computed for each prior according to table (6.1).

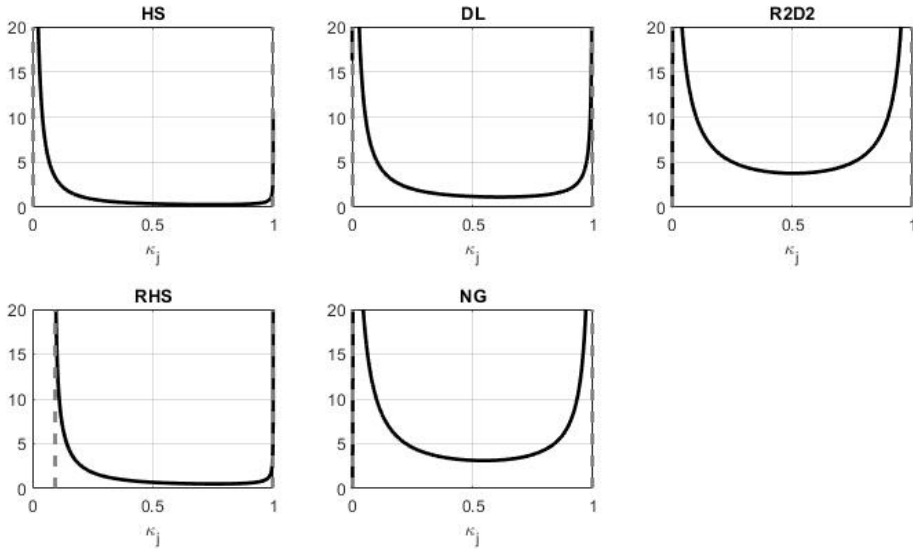


Figure 6.3: Shrinkage Distributions for the HS, DL, R2D2, RHS and NG prior respectively. For the RHS prior, we set $\kappa_0 = 1$. a_j is set to 10 which corresponds to weak global shrinkage, thereby favouring dense solutions. a_j is computed for each prior according to table (6.1).

6.2.4 Prior Predictive Model Size Distributions

While the implied shrinkage coefficient distributions have been derived conditional on the global shrinkage parameter, further information can be gleaned from integrating out the uncertainty over the global shrinkage parameter through prior predictive distributions of the effective model sizes. To do so, we generate global and local scales according to their respective distributions and transform them to κ_j , using to table (6.1), setting $\sigma = 1$ and $T = 100$. To investigate the impact of the size of the covariate vector, we generate these for $K \in \{10, 100, 1000\}$. We set the prior guess for the regularised horseshoe prior to $\kappa_0 = \frac{K}{2}$.

The prior predictive distributions in figures (6.4)-(6.6) show clearly that all but the regularised horseshoe prior with $\vartheta_0 = K/2$ and the A-RHS prior are very informative about the preferred prior effective model size. As expected, by applying a flat Beta prior on the expected ratio of active coefficients, we obtain a uniform prior implicit model size distribution. The R2D2, on the other hand, favours small model sizes compared to the dimensionality of β and the NG tends to form symmetrically around moderately sparse model sizes as the dimensionality increases. The DL along with the HS prior favour very large model sizes. The latter in fact replicates the results from Piironen et al. (2017): the default HS prior is a dubious choice when the interpretation about the degree of sparsity is a priority.

6.2.5 Sparsification Methods

While the presented GL priors typically entail aggressive shrinkage toward zero for noise variables, they do not provide truly sparse posteriors with positive probability, unlike two component mixture priors such as spike-and-slab priors (George and McCulloch, 1993). To conduct model selection which may also be used to draw inference on the degree of sparsity, the literature has proposed many different solutions based on the the coverage rate of the marginal posterior distribution (see Gelman et al. (2013)), point-fit heuristics (Ray and Bhattacharya, 2018; Hahn and Carvalho, 2015; Huber et al., 2019) or fully Bayesian decision analysis (Piironen et al., 2020; Woody et al., 2020)⁶. While relying on coverage rates of the marginal posteriors has appeal from a frequentist perspective, their reliability in detecting signals from noise is largely determined by the degree of collinearity in the data. If groups, or all of the covariates are highly correlated, GL priors tend to shrink subsets of these covariates toward zero in some iterations, but leave them un-shrunk in other iterations. Hence, marginalising over the uncertainty of shrinkage induced

⁶ The methods proposed in Hahn and Carvalho (2015) and related studies use methods from Bayesian decision analysis, however, in comparison to Piironen et al. (2020) can be seen as projecting only the first moment of the predictive distribution of the reference model, while the latter project the entire predictive distribution.

Prior	κ_j	$p(\kappa_j \vartheta, \sigma^2)$
horseshoe	$\frac{1}{1+Ts_j^2\sigma^{-2}\vartheta^2\varphi_j^2}$	$\frac{1}{\pi} \frac{a_j}{(a_j^2-1)\kappa_j} \frac{1}{\sqrt{\kappa_j}} \sqrt{1-\kappa_j}$
regularised horseshoe	$\frac{1}{1+Ts_j^2\sigma^{-2}\vartheta^2\tilde{\varphi}_j^2}, \tilde{\varphi}_j^2 = \frac{c^2}{Ts_j^2+c^2+\vartheta^2\varphi^2}$	$\frac{b-1}{2a(\kappa_j-b)^2 \sqrt{\frac{1-\kappa_j}{a\kappa_j-ab}} \left(\frac{1-\kappa_j}{a\kappa_j-ab} + 1\right)}$
agnostic horseshoe	$\frac{1}{1+Ts_j^2\sigma^{-2}\vartheta^2\tilde{\varphi}_j^2}, \tilde{\varphi}_j^2 = \frac{c^2}{Ts_j^2+c^2+\vartheta^2\varphi^2}$	$\frac{b-1}{2a(\kappa_j-b)^2 \sqrt{\frac{1-\kappa_j}{a\kappa_j-ab}} \left(\frac{1-\kappa_j}{a\kappa_j-ab} + 1\right)}$
dirichlet-laplace	$\frac{1}{1+T\sigma^{-2}s_j^2\psi_j\phi_j^2\vartheta^2}, a_j = \sigma^{-1}s_j\sqrt{T}\vartheta$	$\frac{1/2^{d+1}}{\Gamma(d)} a_j^{-d} \frac{1}{\kappa_j^2} \left(\sqrt{\frac{1-\kappa_j}{\kappa_j}}\right)^{d-2} \exp\left(-\frac{\sqrt{1-\kappa_j}}{\sqrt{\kappa_j}} \frac{1}{2a_j}\right)$
r2d2	$\frac{1}{1+T\sigma^{-2}s_j^2\psi_j\phi_j\vartheta/2}, a_j = T\sigma^{-2}s_j\psi_j$	$\frac{\xi^d}{\Gamma(d)} a_j^{-d} \frac{1}{\kappa_j^2} \left(\frac{1-\kappa_j}{\kappa_j}\right)^{d-1} \exp\left(-\xi \frac{1-\kappa_j}{\kappa_j} \frac{1}{a_j}\right)$
normal-gamma	$\frac{1}{1+T\sigma^{-2}s_j^2\vartheta^{-2}\varphi_j^2}, a_j = T\sigma^{-2}\vartheta^{-2}s_j^2$	$\frac{\theta^\theta}{\Gamma(\theta)} \frac{1}{\kappa^2} \left(\frac{1-\kappa_j}{\kappa_j}\right)^{\theta-1} \exp\left(-\theta \frac{1-\kappa_j}{\kappa_j} \frac{1}{a_j}\right)$

Table 6.1: Shrinkage coefficient distributions implied by the HS, RHS,A-RHS,DL,R2D2 and NG prior respectively. For the regularised horseshoe, we force $b = \frac{1}{1+T\sigma^{-2}s_j^2c^2}$ which entails a Gaussian slab. $\tilde{\varphi}_j$ and a_j can vary according to the exact prior formulation and are therefore supplied seperately in the table. For the R2D2 prior, we follow the recommendations of Zhang et al. (2020), where d is set to $1/(K^{(0.5/2)} \cdot T^{(0.5/2)} \cdot \log(T))$. We set $K = 100$.

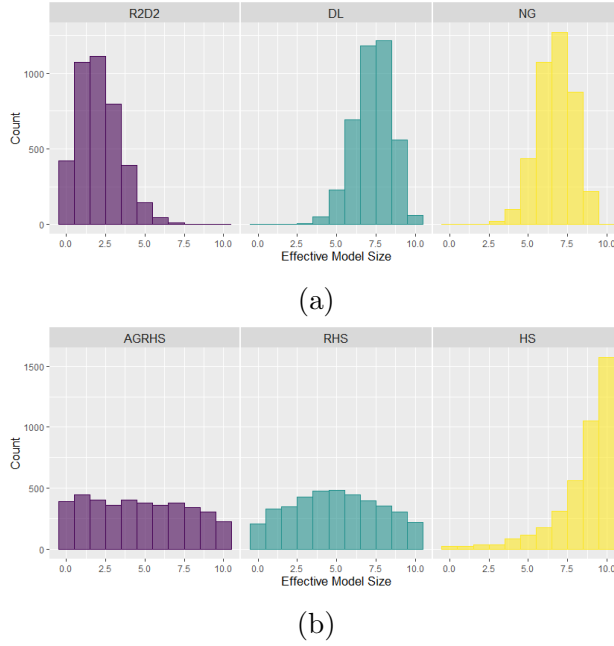
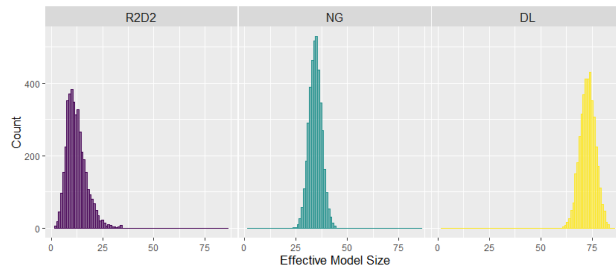
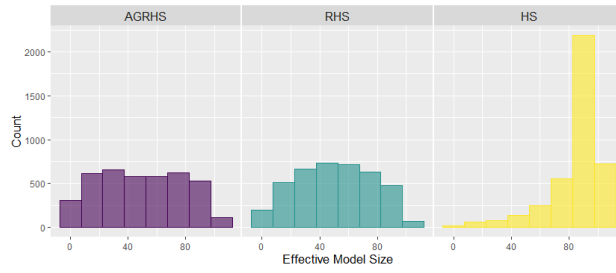


Figure 6.4: Prior predictive model size distributions for $K = 10$.

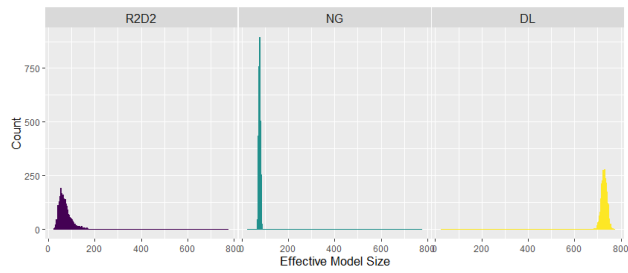


(a)

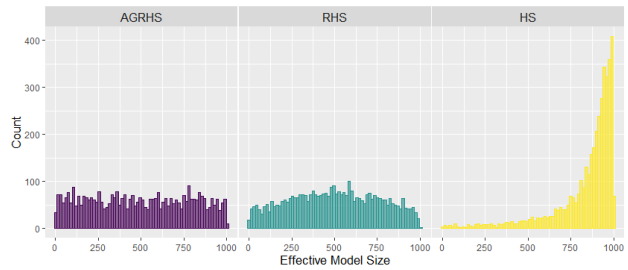


(b)

Figure 6.5: Prior predictive model size distributions for $K = 100$.



(a)



(b)

Figure 6.6: Prior predictive model size distributions for $K = 1000$.

by correlation causes their credible intervals overlapping with zero (Piironen et al., 2020), and can further induce bad mixing due to multimodality⁷. Interpretation on the significance of a single covariate based on the marginal posterior and therefore the posterior model size can therefore be misleading.

Instead, Hahn and Carvalho (2015) and Ray and Bhattacharya (2018) have proposed to threshold variables' posterior coefficient estimates to zero based on heuristics inspired by penalised regression. The SAVS algorithm by Ray and Bhattacharya (2018) is particularly elegant in that it thresholds posterior regression coefficients to zero by:

$$\psi_j = \text{sign}(\bar{\beta}_j) \|X_j\|^{-2} \max(|\bar{\beta}_j| \|X_j\| - \zeta_j, 0), \quad \forall j \in \{1, \dots, K\}, \quad (6.17)$$

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 $\bar{\beta}$ represents a draw from the regression posterior. The parameter ζ_j in (6.17) 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 a simpler solution,

$$\zeta_j = \frac{1}{|\beta_j|^2}, \quad (6.18)$$

which ranks the given coefficient inverse-squared proportionally and provides good performance compared to alternative penalty levels (Ray and Bhattacharya, 2018; Huber et al., 2019). To see the similarity to lasso style regularisation, the solution to (6.17) can be obtained by the following minimisation problem which reminds of Zou (2006):

$$\bar{\psi} = \underset{\psi}{\text{argmin}} \left\{ \frac{1}{2} \|X\hat{\beta} - X\psi\|^2 + \sum_{j=1}^K \zeta_j |\psi_j| \right\} \quad (6.19)$$

$\bar{\psi}$ is the sparsified regression vector. The relative frequency of non-zero entries in the posterior coefficient vector can analogously to the spike-and-slab posterior be interpreted as posterior inclusion probabilities. And integrating over the uncertainty of the parameters to obtain the predictive distribution $p(\tilde{y}|y)$, we receive something similar to a Bayesian Model Averaged (BMA) posterior (Huber et al., 2019; Woody et al., 2020).

Piironen et al. (2020), based on the seminal contributions of Goutis and Robert (1998) and Dupuis and Robert (2003), propose instead to select variables based on the performance of the entire predictive density which minimises the Kullback-

⁷ This of course depends on the flexibility of the sampling algorithm to detect multiple modes in the posterior, which Gibbs and Hamiltonian Monte Carlo methods are known to struggle with (Piironen et al., 2020)

Leibler distance to a reference model containing the full information set of the analyst. In our setting, the reference model will pertain to the posterior predictive distribution using continuous global-local priors. Piironen et al. (2020) name their approach projective predictive inference (projpred) which they show can outperform state of the art frequentist and Bayesian priors for forecasting (Piironen and Vehtari, 2017), but also offers good asymptotic as well as finite sample performance in uncovering the degree of sparsity in the data (Pavone et al., 2020; Afrabandpey et al., 2019). Letting \mathcal{D} stand for the training data, \tilde{y} stand for latent future measurements of y and θ_* be our parameter vector of our reference model, the general idea of posterior projection is to replace the posterior $p(\theta_*|D)$ of the reference model with a simpler posterior $q_{\perp}(\theta)$. $q_{\perp}(\theta)$ might contain user specified preferences such as sparsity, possibly on a different feature space. The domain therefore of $\theta \in \Theta$ might be entirely different to that of the reference model $\theta_* \in \Theta_*$. This prompts not to project via the discrepancy in $p(\theta_*|D)$ to $q_{\perp}(\theta)$, but to project via the predictive distribution which reverts the order of integration:

$$\begin{aligned}
KL(p(\tilde{y}|D)||q(\tilde{y})) &= E_{\tilde{y}}(\log p(\tilde{y}|D) - \log q(\tilde{y})) \\
&= -E_{\tilde{y}}(\log q(\tilde{y})) + \text{const.} \\
&= -E_{\tilde{y}}(\log E_{\theta}(p(\tilde{y}|\theta))) + \text{const.} \\
&= -E_{\theta_*}(E_{\tilde{y}|\theta_*}(\log E_{\theta}(p(\tilde{y}|\theta)))) + \text{const.}
\end{aligned} \tag{6.20}$$

Where $E_{\theta_*}(\cdot)$, $E_{\tilde{y}|\theta}$ and $E_{\theta}(\cdot)$ denote expectations over $p(\theta_*|D)$, $p(\tilde{y}|\theta_*)$ and $q_{\perp}(\theta)$ respectively. KL stands for the Kullback-Leibler distance. Optimal projection of $p(\theta_*|D)$ from Θ_* to Θ is the distribution $q_{\perp}(\theta)$ that minimises the functional (6.20). While minimising the functional in equation (6.20) is often impossible due to the multiple integration steps needed, Piironen et al. (2020) propose a simple and efficient way of conducting projection based on the simulated parameters of $p(\theta_*|y)$. In this approach the Monte Carlo sample $\{\theta_*\}_{m=1}^M$ of the reference model is split into C clusters $\{\theta_*^m : m \in I_c\}$ for $c = 1, \dots, C$ which are then mapped through point estimates for each c into corresponding θ_{\perp} , $\{\theta_{\perp,c}\}_{c=1}^C$ according to:

$$\theta_{\perp} = \operatorname{argmax}_{\theta \in \Theta} \frac{1}{T} \sum_{i=1}^T E_{\tilde{y}_i|I_c}(\log(p(\tilde{y}_i|\theta))), \tag{6.21}$$

where the expectation in equation (6.21) is with respect to the c^{th} cluster. The predictive density can thus be computed by

$$q(\tilde{y}) = \sum_{c=1}^C w_c p(\tilde{y}|\theta_{\perp,c}), \tag{6.22}$$

where the weights are computed proportional to the number of draws in each cluster. Naturally, the more clusters, the higher the accuracy in projecting the reference

posterior but the more computationally intensive the procedure becomes with the amount of Monte Carlo draws; which is usually large for Bayesian models. Conducting posterior projection instead on C clusters provides a good compromise between a variety of simpler and more complicated projection methods (Piironen et al., 2020). For generalised linear models such as in (6.2), whose observation model is in the exponential family, the projection solution $\theta_{\perp,c} = (\beta_{\perp,c}, \sigma_{\perp,c})$ can be derived analytically for each cluster:

$$\beta_{\perp,c} = (X^T X)^{-1} X^T \mu_*^c, \quad (6.23)$$

where $\mu_*^c = \frac{1}{I_c} \sum_{m \in I_c} Z \beta_*^m$ refers to the prediction of the c^{th} cluster based on feature set Z of the prediction model which, as mentioned, can be different to the features used for the submodels, X. The noise variance for each submodel is similarly derived as:

$$\sigma_{\perp,c}^2 = \frac{1}{T} \sum_{i=1}^T V_i^c + \frac{1}{T} \|X \beta_{\perp,c} - \mu_*^c\|^2, \quad (6.24)$$

where V_i^c is the predictive variance of \tilde{y}_i from the reference model, for cluster c. This can be derived from the law of iterated expectations:

$$\begin{aligned} V_i^c &= Var(\tilde{y}_i | I_c) = E(Var(\tilde{y}_i | \beta_*, \sigma_*^2) | I_c) + Var(E(\tilde{y}_i | \beta_*, \sigma_*) | I_c) \\ &= \frac{1}{|I_c|} \sum_{s \in I_c} \sigma_{*,m}^2 + V_{s \in I_c} [z_i^T \beta_*^m]. \end{aligned} \quad (6.25)$$

Note the intuition in equation (6.24): the projected variance parameter is a linear combination between the variance of the reference model and the difference between the reference model predictions and the sub-model predictions. This entails that the predictive variance of the projected model has the predictive variance of the reference model as a lower bound. Similarly, projection onto the reference model acts a safeguard to overfitting: instead of potentially overfitting noise in the actual data, overfitting to the reference model might still produce good out-of-sample performance on the actual data.

Finally, although projection offers a safeguard to overfitting, here we are interested in inferring about the degree of sparsity in the data. This motivates, as in Piironen et al. (2020), to use a search heuristic to find the optimal, possibly sparse combination of covariates, that come as close as possible or even improve the predictive performance of the reference model⁸. Similar to the SAVS approach, Piironen et al. (2020) propose L_1 -penalisation akin to adaptive lasso:

$$\min_{\beta} \left\{ -\frac{1}{T} \sum E_{\tilde{y}_i} (\mathcal{L}_i(\beta, \tilde{y}_i)) + \vartheta \sum_{j=1}^K \varphi_j \left(\frac{1}{2} |\beta_j| \right) \right\} \quad (6.26)$$

⁸ Piironen et al. (2020) lay out the conditions under which Projective predictive inference can improve on the predictions of the reference model.

We follow Piironen et al. (2020) by validating ϑ through Pareto-smoothed importance sampling leave-one-out (PSIS-LOO). The PSIS-LOO method provides an approximation to the true LOO performance based on an importance sampling procedure which accounts for possible degeneracy in the importance weights (Vehtari et al., 2017). PSIS-LOO has the advantage that the Bayesian reference model need only be computed once instead of T times for true LOO predictive checking. Vehtari et al. (2017) have established that PSIS-LOO not only has asymptotic validity but performs better in finite samples than other popular information criteria based on predictive performance such as the WAIC, AIC and DIC. To automate decision making about the model sizes implied by choosing the value of ϑ , we follow Piironen et al. (2020) by selecting the smallest model size k that has a PSIS-LOO expected log-predictive density score (elpd) no more than one standard error away from that of the reference model (see section 5.2 in Piironen et al. (2020)).

Note that compared to SAVS, projpred is not only based on out-of-sample predictive performance approximated by Bayesian LOO, but also offers a fully Bayesian approach to variable selection in the sense that the entire reference posterior predictive distribution is projected onto sub-set models, whereas the SAVS approach only projects the first moment of the in-sample predictive distribution. Another difference is that projpred’s projection is optimal in the sense of a Kullback-Leibler loss, whereas SAVS is optimal under squared loss. It is interesting to note that for the Gaussian linear regression, the projection of the regression parameters are identical.

One of the main contributions of this chapter is to investigate whether the sparsification through SAVS or projective predictive inference is able to uncover the degree of sparsity similarly to spike-and-slab prior as well as whether these sparsification techniques improve on forecasting performance.

6.3 Simulation

In this section, we investigate the methods in capturing the degree of sparsity in the data as well as how well they forecast on simulated data sets. We generate 100 fictitious data sets according to the linear regression model (6.2), where β either follows a sparse or dense representation. In particular, we generate $T = 150$ observations for model (6.2), with $K = 100$, and X drawn from a multivariate Gaussian with zero mean and Toeplitz covariance matrix, Σ , where $\Sigma_{i,j} = \rho^{|i-j|}$ and $\rho = 0.5$. The true coefficient vectors represent a sparse and dense setting, where β^{sparse} has 5 non-zero entries and β^{dense} has 80. The non-zero coefficients are drawn from a normal distribution:

$$\beta_j^s \sim N(0, 0.5^2), \quad j \in \{1, \dots, K\}, \quad s \in \{sparse, dense\}. \quad (6.27)$$

The coefficient vectors are additionally populated randomly. For all simulations throughout, we define $\epsilon \sim N(0, \sigma^2)$ with $\sigma = 1$.

For each prior in table (6.1)⁹, we sample the implicit posterior model size distribution according to equation (6.6), based on the prior specific transformation as in the second column of table (6.1). As a benchmark for the proposed A-RHS model, we set ϑ_0 for the RHS prior to the best guess possible, equal to the correct amount of active coefficients ($\vartheta_0 = 5$, or $\vartheta_0 = 80$). We would expect that if the data are informative about the degree of sparsity, then the agnostic RHS can only do as well as the RHS with the correct guess. Model size distributions are computed for the SAVS algorithm as the number of non-zero coefficients of the sparsified posterior regression draws for each prior respectively, and for the projpred approach, we use the suggested model size methodology as described in section 6.2.5. And finally, to provide a benchmark on the effect of generalising the slab distribution of the SSVS prior, we also estimate the t-distributed SSVS prior of Fava and Lopes (2021) with 3 degrees of freedom¹⁰¹¹.

For each method, we report in table (6.2) the average of the model size distributions (model point estimates in case of projpred) across all 100 data sets.

Forecasts are generated from the predictive distribution

$$p(\tilde{y}|y) = \sum_{m=1}^M p(y|\Upsilon)p(\Upsilon|y), \quad (6.28)$$

where Υ contains samples of $(\beta, \beta^{SAVS}, \phi_{\perp}, \sigma^2, \sigma_{\perp}^2)$, depending on the method used^{12 13}.

Forecasts are evaluated according to root-mean-squared forecast errors (RMSFE) based on the last 50 observations which are not used to compute the posteriors. For point-forecasts \hat{y} , we use the squared loss optimal mean of the predictive distribution:

$$RSMFE = \sqrt{\frac{1}{50} \|\hat{y} - \tilde{y}\|_2^2} \quad (6.29)$$

⁹ The NG prior tended to fair worse than the other priors in our simulations and was left out for parsimony. Results available upon request.

¹⁰ The results were not sensitive to the degrees of freedom below the threshold of 20, after which the t-distributions becomes very close to the normal.

¹¹ Code for this model was provided by the authors.

¹² For the SAVS sampling approach, we sample β^{SAVS} and do not resample σ^2 after sparsification. We found that there is no practical difference in the SAVS' performance regardless of resampling.

¹³ For projpred, we stick to the recommendations of the R package and use 5 clusters for forecasting.

Forecast density fit is measured by the continuous rank probability score (CRPS):

$$\text{CRPS} = \frac{1}{50} \sum_{t=1}^{50} \frac{1}{2} \left| y_{100+t} - \hat{y}_{100+t} \right| - \frac{1}{2} \left| \tilde{y}_{100+t}^A - \tilde{y}_{100+t}^B \right|, \quad (6.30)$$

where $\tilde{y}_{100+t}^{A,B}$ are independently drawn from the posterior predictive density $p(\tilde{y}|y)$. The CRPS belongs to the class of strictly proper scoring rules (Gneiting and Raftery, 2007), and can be thought of as the probabilistic generalisation of the mean-absolute-forecast-error¹⁴.

Table 6.2: Simulation Evaluation Metrics

Priors	(I)MS	SAVS	Proj.	Unspars.	SAVS	Proj.	Unspars.	SAVS	Proj.
Sparse DGP									
	Model Sizes			RMSFE			CRPS		
HS	11.51	2.86	5.4	1.09	1.09	1.10	0.22	0.22	0.22
RHS	7.17	2.82	5.45	1.09	1.09	1.1	0.23	0.22	0.23
A-RHS	9.41	3.02	5.55	1.09	1.09	1.1	0.22	0.21	0.23
DL	55.69	12.22	1.2	1.24	1.12	1.11	0.25	0.23	0.23
R2D2	11.56	3.22	2	1.09	1.09	1.1	0.22	0.21	0.22
SSVS	13.51			1.1			0.23		
SSVS-t	4.62			1.19			0.24		
Dense DGP									
	Model Sizes			RMSFE			CRPS		
HS	62.52	43.47	70.85	2.25	2.38	2.37	0.45	0.48	0.46
RHS	80.83	45.69	71.05	2.21	2.21	2.26	0.45	0.47	0.46
A-RHS	76.14	48.56	75.9	2.24	2.24	2.27	0.47	0.45	0.47
DL	55.8	42.67	79.85	2.22	2.32	2.27	0.48	0.47	0.45
R2D2	18.3	24.79	74.9	2.9	2.82	2.71	0.67	0.59	0.57
SSVS	75.33			2.24			0.47		
SSVS-t	68.38			2.32			0.49		

The first vertical panels (Model Sizes) show the average over simulations for implicit model sizes from the global-local priors as well as those from the SSVS priors (column IMS), from the SAVS algorithm (column SAVS), and the projective approach (Proj.). RMSFE and CRPS are computed for un-sparsified, SAVS and projective inference sparsified approaches in column pertaining to RMSFE and CRPS respectively. The upper panel shows the results from the sparse DGP, while the lower one shows those from the dense DGPs. The RHS prior is estimated with the correct ratio of active to inactive coefficients.

¹⁴ Results using the log-score (Gneiting and Raftery, 2007) were very similar and available upon request.

The results indicate, as expected, that on average the implied model sizes centre on sparse and dense models for priors that accommodate sparsity and density naturally such as the horseshoe prior variants and offer competitive performance to the spike-and-slab priors. In fact, the average implicit model sizes based on the RHS and A-RHS are closer to the truth than the spike-and-slab priors for the dense DGP, and are also slightly closer to the true model size in the sparse DGPs compared to the SSVS prior. In line with our discussion, the RHS prior with the correct active coefficients guess, shows the overall best IMS point estimates among the continuous shrinkage priors. However, the agnostic-RHS already comes very close to that performance despite remaining relatively uninformative about implicit model sizes a-priori. Both the RHS and A-RHS beat out the HS models in terms of implied model sizes, as expected.

The implied model sizes for the remaining GL priors showcase behaviour that mimics the behavior found in the prior predictive distributions. The R2D2 prior over-shrinks, likely due to the default prior hyperparameter choice which results in particularly poor performance in the dense setting. The DL prior seems to favour a medium level of sparsity, independent of the true sparsity in the DGP. Using further hyper-priors on the sparsity controlling parameters of these priors might be able to make change this behaviour which we suggest as a fruitful avenue for future research.

The sparsification approaches broadly confirm the findings of the implicit model size estimates, although the general pattern that emerges is that the SAVS over-sparsifies and that projective inference can remedy some of the short-comings of the priors in terms of pinning down the correct model sizes. How accurate the model sizes are, however, is determined by the quality of shrinkage of the priors. As such, the SAVS shows too large model sizes for the DL prior in the sparse DGP which the IMS point estimates indicate already under-shrinks, and too small model sizes for the R2D2 in the dense DGP, also in line with its IMS estimate. Looking at the projpred approach, the picture changes. Here we find that it can vastly improve the sparsity estimates even for the DL and R2D2 prior, particularly in the dense DGPs. This is likely explained by projpred taking into account out-of-sample predictive performance as measured through the approximate PSIS-LOO which is more heavily influenced in the dense DGP by inclusion of the active covariates.

Forecasting performance is proportional to the priors' as well as sparsification technique's ability to pin down the degree of sparsity. Priors with poor prior implicit model size and sparsification performance, such as the DL (sparse and dense) and R2D2 (dense), produce the worst forecasting results. As expected, the forecasting performance of SAVS and projpred is mostly preserved compared to the un-sparsified posteriors, where projpred produces the largest gains in forecasting per-

formance, when the prior either under shrinks (DL, sparse) or over shrinks (R2D2, dense). While the selection of the global-local priors produces competitive performance compared to the spike-and-slab priors, the RHS prior tends to slightly outperform. Finally, it is interesting to note, that although model size identification (with the caveats above) tend to do similarly well in sparse and dense settings, forecasting performance is markedly worse for all priors the dense simulations. This is in line with previous research such as Cross et al. (2020); Kohns and Bhattacharjee (2022).

The simulations therefore offer evidence that the global-local prior’s ability to detect the degree of sparsity in the data is sensitive to the choice of the specific global-local prior and their hyperparameters. Using the R2D2 prior with it’s default settings for example results in overshrinkage, especially in dense DGPs, which might give a false indication through IMS, SAVS or Projpred, of sparsity (“illusion of sparsity”). However, with appropriately chosen priors such as the RHS and A-RHS, one can expect not only good forecasting performance but also good sparsity detection which is competitive with the spike-and-slab. Of the sparsification approaches, projective inference tends to produce the most reliable estimates about sparsity without harming predictive performance too much. The simulations also underscore that pinning down the degree of sparsity is not only important from an interpretability standpoint but also determines how well the model is predicting out-of-sample.

6.4 Application to Economic Data Sets

For empirical validation of the simulation results in section (6.3), we now set out to compare the same set of priors in terms of predictive accuracy (sparsified and non-sparsified) and sparsity detection by applying them to a variety of popular ‘Big Data’ sets which have received much attention in the economic literature. To increase comparability to Giannone et al. (2021), we use the same data sets. These comprise:

- Macro 1: investigates predictive fit of a large macro data base, originally employed by Stock and Watson (2002) for forecasting monthly US industrial production. This data set contains 130 possible predictors and the sample spans 659 monthly observations from February 1960 to December 2014
- Macro 2: investigates cross-country determinants of growth, based on Barro (1991). Comprises 60 economic variables to predict average growth rate of GDP over the sample 1960-1985 with in total 90 observations
- Finance 1: investigates predictive ability of 16 lagged macro indicators in forecasting annual S&P 500 returns. Contains 68 annual observations between 1948 and 2015. Based on Welch and Goyal (2008)
- Micro 1: investigates the effects of a abortions on crime rates in the US, controlling for 284 covariates. Contains 576 panel observations for US states between 1986 and 1997. Based on Donohue III and Levitt (2001)
- Micro 2: Estimates the determinants of eminent domain. The dependent variable is pro-plaintiff eminent domain decisions which is regressed on 138 characteristics and interactions of judicial panels. Includes 312 panel observations between 1979-2004. Based on Belloni et al. (2012)

More information on the economic settings and purpose of these studies can be found in Giannone et al. (2021) and references therein.

Similar to the simulation study above, we investigate the degree of sparsity from our selection of global local priors through (1) implicit model sizes (IMS), (2) SAVS sparsification, (3) projective predictive inference for which we report average model sizes as point estimates in table (6.3), histograms of posterior model sizes in figures (E.1-E.10) and coefficient inclusion heatmaps in figures (E.11-E.20) as in Giannone et al. (2021). For all of the displayed results, we split the data sets in half ¹⁵, where the first tranche is used for model size computations and the latter tranche is used for forecast evaluation. Again, we compare forecasts via RMSFE and generate

¹⁵ Except for Micro1, as the SSVS priors were not able to converge with exactly splitting the sample in half. Instead, we hold out a sample of the latter 250 observations for forecast evaluation.

the point forecasts from the respective predictive distributions using the hold-out sample¹⁶. The priors are implemented as outlined in section 6.2.2. The RHS is implemented with the default setting recommended by the Stan community when no information about the number of active coefficients is available, (Carpenter et al., 2017), $\kappa_0/(K - \kappa_0) = 1$.

Priors	Posterior Model Sizes														
	Macro 1			Macro 2			Finance 1			Micro 1			Micro 2		
	IMS	SAVS	Proj	IMS	SAVS	Proj	IMS	SAVS	Proj	IMS	SAVS	Proj	IMS	SAVS	Proj
HS	18.14	2.83	6	7.6	1.17	3	9.14	3.13	1	0	0	2	52.5	10.17	7
RHS	19.24	3.23	2	5.54	0.84	3	8.88	2.77	2	3.91	0.06	14	65.89	17.84	138
A-RHS	17.06	3	6	3.18	0.41	3	6.85	2.1	2	3.22	0.05	14	22.55	4.01	7
DL	74.17	19.66	6	36.63	13.07	3	9.42	3.27	2	197.97	53.38	14	105.034	39.7	138
R2D2	14.94	3.04	6	6.91	1.13	3	5.487	1.98	2	17.49	0.09	14	17.41	2.51	7
SSVS	29.68			24.55			11.34			11.06			88.4421		
SSVS-t	5.59			4.43			7.45			2.88			30.06		

Table 6.3: Expected posterior model sizes based on implicit model sizes, the SAVS algorithm and projective inference. Posterior model sizes for the SSVS and SSVS-t are based on posterior inclusion.

Priors	RSMFE														
	Macro 1			Macro 2			Finance 1			Micro 1			Micro 2		
	IMS	SAVS	Proj	IMS	SAVS	Proj	IMS	SAVS	Proj	IMS	SAVS	Proj	IMS	SAVS	Proj
HS	0.75	0.74	0.75	1.02	1.04	1.02	1.33	0.99	1.01	1.25	1.25	1.22	1.12	1.1	1.03
RHS	0.75	0.74	0.78	1.01	1.04	1.03	1.17	0.96	0.95	1.21	1.22	1.21	1.21	1.14	1.18
A-RHS	0.75	0.74	0.76	1.02	1.04	1.03	1.07	0.94	0.94	1.21	1.22	1.21	1.12	1.11	1.11
DL	0.8	0.76	0.77	0.96	0.95	1.03	1.21	1.01	0.95	1.31	1.26	1.24	1.36	1.27	1.31
R2D2	0.75	0.74	0.77	1.01	1.03	1.03	1.02	0.94	0.93	1.21	1.22	1.21	1.11	1.11	1.11
SSVS	0.76			1.04			1.12			1.21			1.16		
SSVS-t	0.86			1.01			1.3			1.21			2.46		

Table 6.4: Root-mean-squared forecast error (RMSFE) for the plain priors, SAVS and projected posteriors. RMSFE's are computed by generating the predictive posterior distribution $p(\tilde{y}|y)$ based on the out-of-sample observations.

The first point to notice from table (6.3) is that the average degree of sparsity, compared to the dimensionality of the data, across most GL priors¹⁷ is similar to the SSVS for all data sets but Micro 2 and Macro 2 as measured by the posterior implicit model sizes. These averages, however, hide the vastly different model size uncertainty between the GL and SSVS priors. Plotting the distributions of IMS next to the SSVS in figures (E.1-E.5), reveals that even for the data sets where the average posterior model size is similar, the SSVS displays far greater model size uncertainty and disagrees starkly in terms of the modes for Finance 1 and Micro 2. These results agree with findings of Fava and Lopes (2021) who show that the parametric assumption about the normality of the slab distribution of the SSVS, may impose too strong shrinkage, therefore inflating model size uncertainty. In fact, by increasing the flexibility of the slab distribution by allowing for a t-distribution (SSVS-t) with three degrees of freedom, model-size distributions become more similar to those of the GL

¹⁶ Density evaluation results mirror the performance of RMSFE, so are left out for parsimony. Available upon request.

¹⁷ Similar to the simulations, the DL prior favours dense implicit models.

priors, particularly horseshoe priors. That said, Finance 1 and Micro 2 appear to be challenging data sets for all priors, as can be seen from the heterogeneous forecasting results and obvious convergence problems, particularly for the HS and RHS; for which implicit model size distributions are bi-modal. This is also confirmed by more divergences in the Hamiltonian Monte Carlo implementation of the RHS compared to the A-RHS. While the RHS produces more than 30 divergences per chain for the Micro 2 data set, the A-RHS produces none. However, those GL priors that offer good convergence as gauged by the IMS distributions, such as the A-RHS and R2D2, far outperform the other GL and SSVS priors in terms of forecasting. In fact, the A-RHS and R2D2 are the only priors that consistently outperform the SSVS prior in terms of forecasting. Taking the findings from model size distributions and forecast accuracy together, it may be concluded that the SSVS’ parametric assumptions are too strict for challenging data sets such as Finance 1 and Micro 2. Forecasting results for the remaining Macro 1, Macro 2 and Micro 1 data sets confirm the similarity between all priors with nearly identical RMSFE.

Similar to the simulation results, applying SAVS to the GL regression posteriors acts to strongly sparsify, with model sizes consistently lower than IMS, while keeping forecasting performance either stable or increasing it. Performance increases are especially noticeable for Finance 1 and Micro 2. The reason is simple: comparing the covariate inclusion heatmaps between GL priors and the SSVS, SAVS dramatically reduces model uncertainty by identifying the covariates driving fit, which is especially evident for Finance 1 (E.13). Heatmaps for the other data sets show that only in Macro 1, is the SAVS procedure able to identify dominant covariates. For Macro 2 (E.12) and Micro 1 (E.14), in contrast, virtually all coefficients are shrunk to zero with no dominant regressor, despite nearly identical forecasting performance to the non-sparsified priors. This shows that predictability is very low for these data sets; which is information that the plain priors are not able to communicate without post-estimation sparsification. While these results are encouraging for the SAVS, benefits of its application are limited by the quality of shrinkage and convergence of the prior: convergence issues for the RHS carry over to the SAVS model size distribution in Micro 2 which also results in worse forecasting performance compared to the other priors, and, in coherence with the simulations, SAVS model sizes for the DL prior are consistently larger than the other GL priors, indicating less aggressive shrinkage.

Projective predictive inference, similar to the SAVS procedure, is able to either retain forecasting performance or even marginally outperform both the plain prior and SAVS. Clear examples of the latter can be found for the DL prior in Finance 1 and the HS prior in Micro 2. A common feature in these situations is that the priors do not perform adequate sparsification, even with the SAVS algorithm, which projective predictive inference can partly remedy due to variable selection based on

out-of-sample performance. However, in situations where the priors themselves do not identify signals from noise, such as for the DL prior in Micro 2, projpred fails to improve forecasting performance or indicate a realistic model size. The lacklustre shrinkage induces huge uncertainty in the elpd scores, as can be seen in figures (E.21)-(E.22), which leads to obscure suggested model sizes by the decision rule used (3.5). In general, as for SAVS sparsification, Projpred usually helps predictive performance, however can be limited by the quality of the prior’s shrinkage.

The empirical applications have confirmed the intuition gleaned from the Monte Carlo simulations. Not all global local priors are alike. Flexible priors such as the RHS and A-RHS are competitive at forecasting with the gold-standard of Bayesian variable selection priors and can produce similar measures of sparsity through implicit model sizes. These can be more robust in challenging data sets due to less restrictive parametric assumptions as well as the availability of theoretically sound convergence diagnostics such as Hamiltonian divergences¹⁸. Sparsification based on out-of-sample predictive performance, is reliable at identifying the best subset of signal variables which either retains predictive accuracy, or increases predictive performance whenever the prior has correctly identified signals and noise. It is the last point that makes projective predictive inference a particularly valuable tool for the types of regressions considered in this chapter, namely to investigate the degree of sparsity with the joint goal of good predictive accuracy. Nevertheless, the presented measures of sparsity may be best used in conjunction to communicate on the one hand a model-internal perspective on sparsity as well as a forecasting perspective that finds the best subset for prediction via projection.

However, care should be taken in interpreting the output of sparsity measuring approaches for global-local priors which embed strong prior beliefs about sparsity such as the R2D2 and DL priors with their default settings.

6.5 Conclusion

In this chapter we have set out to investigate whether a host of global-local priors are able to inform on the degree of sparsity in economic data sets and whether forecast performance is competitive with default prior for these tasks, the spike-and-slab prior. This endeavor to consider GL priors is motivated by two observations. Firstly, spike-and-slab priors due to their discrete mixture formulation can be sensitive to choices in hyper-parameters and parametric assumptions about the slab distribution have a non-trivial impact on variable selection and prediction (Fava and Lopes, 2021). Secondly, global-local priors tend to have excellent forecasting and theoret-

¹⁸ Discrete mixture priors are hard to implement with the Hamiltonian Monte Carlo method, as the transition probabilities are derived under the assumption of continuous parameter spaces (Carpenter et al., 2017).

ical properties (Bhadra et al., 2019; Zhang et al., 2020), but lack interpretability in terms of communicating the inferred degree of sparsity due to their continuous nature. For this, we analysed the implied degree of sparsity via implied model size distributions, proposed a new horseshoe prior variant which a-priori stays agnostic about the implied degree of sparsity and additionally investigate sparsity from post-estimation sparsification approaches as a comparison. The latter are motivated from Bayesian decision theory, where SAVS finds the optimal posterior subset of covariates via squared error loss, while projective inference finds that subset by projecting the entire posterior that minimises Kullback-Leibler loss. Projective inference additionally takes out-of-sample performance into account via Bayesian leave-one-out cross-validation.

Prior predictive distributions for the implied model sizes show novel results in that all but the agnostic regularised horseshoe prior are informative about the degree of sparsity. Particularly the DL and R2D2 priors are shown to carry over to their tendencies from their prior implied degree of sparsity also to their posterior estimates in simulation and real-economic data sets, irrespective of the true sparsity in the DGP. Priors such as the RHS with correct guesses on sparsity as well as the agnostic RHS provide competitive implicit model sizes compared to spike-and-slab priors while being able to offer similar or slightly better forecasting performance. The agnostic RHS supports findings from Fava and Lopes (2021) that the default SSVS implementation may inflate model size uncertainty likely related to the strict parametric assumptions of the slab distribution.

Sparsification approaches confirm the findings from implicit model size distributions. The quality of shrinkage will heavily influence the ability of either sparsification approach to detect signal from noise. Both simulation and empirical evidence suggest, however, that the extra flexibility of projective inference to use out-of-sample predictive performance for variable selection can improve on the popular SAVS algorithm which tends to over-sparsify, particularly in dense DGPs. For horseshoe based priors, variable selection with projective inference yields competitive estimates of the degree of sparsity compared to the spike-and-slab.

This chapter, therefore gives new impetus to the recent debate on the “illusion of sparsity” in economic data sets. We have shown that flexible priors such as the RHS with adequate sparsity guesses can often outperform the standard SSVS prior for both prediction and estimates of the degree of sparsity, and that our proposed agnostic-RHS prior can automate that guess in a robust data-driven way.

A caveat for this and related studies is that the assumption of a fixed degree of sparsity might be inappropriate for many macroeconomic time-series prediction problems in which changes in the correlation structure are an empirical regularity (McAlinn et al., 2018; Frühwirth-Schnatter and Lopes, 2018; Rockova and McAlinn, 2021) with recessions and large shocks such as the recent Covid-19 pan-

demic (Antolin-Diaz et al., 2021). The recently proposed dynamic horseshoe prior (Kowal et al., 2019) could provide a promising avenue for the study about sparsity over time and is amenable to the methods presented in this chapter. Likewise unexplored in this chapter is the role of grouping induced by large cross-sectional correlation between groups of covariates. Grouping is a natural phenomenon in macro time-series and mixed frequency applications where high-frequency lags and thematically close macro indicators have high correlation. Chapter 2 gives new insights using the GIGG prior, but a more rigorous investigation across grouped shrinkage priors in vain of this chapter would similarly help solidify the impact of grouping on our interpretation of the degree of sparsity.

Chapter 7

Conclusion

Unprecedented availability of commonly employed, but also new data sources, such as Internet search terms, credit card spending or satellite data, for macroeconomic nowcasting and risk estimation necessitate methods that, on the one hand, cope with the high-dimensionality of the estimation problem, but also remain interpretable so as to be communicable to the policymaker. Interpretability, however, is complicated in these fields, among other things, by latent-time trends, complicated correlation structures, severe economic shocks such as the Covid-19 pandemic and heterogeneous variable impact across the conditional distribution of the target variable.

This thesis contributes to the methodology for nowcasting and risk estimation by extending frontier Bayesian priors and computational algorithms to address these problems, while giving interpretability to the respective model predictions using Bayesian decision theory.

In chapter 2, we propose a Bayesian Structural Time Series model for nowcasting quarterly GDP growth. We model slow moving changes in general economic conditions, akin to the frontier of the DFM nowcasting literature (Antolin-Diaz et al., 2017) with latent trends, and extract high-frequency information with a high-dimensional Bayesian MIDAS component. To regularise the highly parameterised model, we use non-centred state space methods (Frühwirth-Schnatter and Wagner, 2010) that facilitate selection of trends and extend the horseshoe prior (Carvalho et al., 2010) for MIDAS. To communicate variable importance for the nowcasts, we apply the SAVS algorithm (Ray and Bhattacharya, 2018) which finds the best subset of mixed frequency variables that summarise model predictions. A novel nowcasting application in which we merge traditional macroeconomic data with a high dimensional Internet search term set, as well as a simulation exercise, show the adaptability of the proposed methods in response to challenging time-series and sparsity structures. We show that the prior and selection algorithm outperform traditionally employed spike-and-slab priors within our model.

In chapter 3, we generalise the BSTS model further with fat-tailed stochastic volatility and a new combination of group-prior and sparsification algorithm. The

former allows the proposed model to flexibly adapt to vastly different time-series dynamics such as those experienced pre- and during the Covid-19 pandemic, and the latter directly addresses correlation clustering issues that arise due to mixed-frequency data transformations. A variable selection algorithm is proposed which follows the decision theoretic paradigm and enables the policymaker to interpret the joint effect of groups of mixed-frequency lags on nowcasts. Accounting for the group-structure helps to avoid mixing issues and random within-group variable selection which may occur with highly correlated mixed frequency data using non-grouped shrinkage priors (Piironen et al., 2020). An application to the UK economy allow to conclude that the proposed methodology vastly improves on competing frontier nowcasting methods, particularly in response to the pandemic. A simulation exercise confirms the favourable performance of prior and variable selection compared to competing methods, especially for dense and highly correlated DGPs.

In chapter 4, we contribute to the growth-at-risk (GaR) literature by extending the global-local shrinkage prior framework, with the horseshoe as our focal point, to the Bayesian quantile regression and provide a new computational algorithm that vastly speeds up estimation in high-dimensional risk estimation problems. The methods are applied to a novel high-dimensional GaR application to the US in which we allow a large set of macro indicators to heterogeneously impact quantiles of the combined conditional GDP distribution. This generalises previous GaR approaches (Adrian et al., 2019) which had a-priori restricted quantiles to only respond to financial conditions. The horseshoe prior BQR is shown to vastly outperform quantile specific and density forecasts from previously proposed continuous shrinkage priors. The good performance of the HS-BQR is further confirmed in a large simulation exercise.

In chapter 5, we derive the Bayesian decision theoretic variable selection approach for the Bayesian quantile regression to harness the forecasting capabilities of continuous shrinkage priors while remaining interpretable to the policy maker. We derive novel analytical solutions and propose a quantile specific parameter for the the algorithm that is data dependent and allows automatic adaptation to the quantile specific degree of sparsity. We use this methodology to extend our high-dimensional growth-at-risk application to the US for which we are now able to communicate drivers of GDP risk across quantiles and across time. Our methods reveal that risks to the US economy vary starkly over time, in line with our understanding of the evolution of macro risks to the US economy. We show that our sparsification can even improve downside risk estimation while not compromising the calibration of the quantiles. The good performance is confirmed in a range of simulations, in which it compares favourably to the spike-and-slab variable selection prior.

In the context of our models, the proposed Bayesian decision theoretically motivated variable selection methods in chapter 2-5 have shown that selection and

prediction can be superior to spike-and-slab priors. Previous work by Giannone et al. (2021) caution that the model informed notion of sparsity can be misleading if the prior (or sparsification method) is *a priori* informative about sparsity. In this chapter, we analyse the general properties of global-local priors to detect the degree of sparsity in the data. We consider two methods of analysis: implicit model size distributions which are derived following Piironen et al. (2017), and sparsification techniques that are tested on a host of economic data sets and simulations. The findings motivate a new horseshoe type prior where we explicitly model the prior on the degree of sparsity to be uniform over any sparsity outcome. The empirical findings from simulations and economic data sets support the theoretical analysis that the proposed prior provides competitive forecasts while the sparsity estimates are often more robust than that of the spike and slab prior.

Admittedly, there are several limitations to the presented work which also serve as fruitful avenues for future research. Methodologically, the priors considered for the presented nowcasting and risk estimation models, have focused on horseshoe or horseshoe like (group-)priors due to their flexibility in allowing for both dense and sparse model representations. The literature on global-local priors, however has been moving fast, and there are now certainly more and perhaps more flexible priors to consider, e.g. those mentioned in Cadonna et al. (2020), which are able to nest many of the priors studied in chapter 6, but also priors that integrate insights from the machine learning literature into their computation and functional representation (Shin and Liu, 2021). Further, for the proposed large dimensional growth-at-risk models, extensions to group-priors would be able to address correlation clustering that is often observed for macro indices pertaining to similar economic sectors (McCracken and Ng, 2016) and has been overlooked so far for the BQR modelling framework. Extending the GIGG prior and group-shrinkage algorithm to the BQR would make for a compelling contribution for macro risk applications.

A commonality faced by both nowcasting and risk estimation models is that predictions can be influenced by different variables over time. This is particularly true in using Big Data sources such as Internet search terms which can display large cyclicity in relevance for predictions, but also risk estimation, where chapter 5 has shown that drivers of risk vary substantially over time. Likewise, chapter 4 has shown that macro indices selected into the cyclical component are different pre- vs. during the pandemic. Hence, a valuable contribution to the literature would be to extend the proposed methods to time-varying shrinkage and sparsity. For this, one might consider the dynamic horseshoe prior (Kowal et al., 2019) for modelling time-varying MIDAS and Bayesian quantile coefficients. A cost associated with using dynamic shrinkage priors, on the other hand, is that the already large parameter spaces such as for the applications chapters 2, 4, 5 and 6 are expanded even further

which poses problems for accurate inference. To induce parsimony in the processes that govern time-variation, a promising alternative approach could be to restrict the joint time-processes to a lower dimensional factor representation such suggested for VARs in Chan et al. (2020). Since it is likely that the parameter processes will co-move for highly correlated data, this approach also aids interpretability in analysing the time-trends.

This thesis has focused on making extensions to priors and variable selection tools for linear or semi-parametric methods that are shown to improve forecast accuracy, but crucially, offer a simplicity that eases communication of the results to the policymaker. Non-parametric methods such as Gaussian processes (Williams and Rasmussen, 2006) and Bayesian artificial regression trees (Chipman et al., 2010), on the other hand, are becoming increasingly popular in social science (Prüser, 2019; Huber et al., 2020; Gramacy, 2020) and offer a new flexibility in modelling complicated non-linearities in time-series methods (Roberts et al., 2013). Within the context of the presented models, non-parametric methods could allow the latent trend and stochastic volatility components to potentially adapt even more flexibly to fast moving non-linearities and uncover co-dependencies in the unobserved components during times of large economic distress, such as the Covid-19 pandemic.

Further, something this thesis stayed largely silent on, are the large sample properties of combining shrinkage priors with sparsification algorithms for variable selection. While Chakraborty et al. (2020), Ray and Bhattacharya (2018) and Woody et al. (2021) provide results on posterior consistency and prediction risk for normal linear regression models with the horseshoe prior in combination with the SAVS (Ray and Bhattacharya, 2018), large sample properties for the algorithms applied to the GIGG prior as well as the Bayesian quantile regression have been left unexplored. It would therefore be a valuable contribution to the literature to investigate the asymptotics for the proposed approaches, particularly in comparison to spike-and-slab priors, so as to give the precise conditions under which the decision theoretically motivated way to obtain sub-set selection is preferable (if at all).

Similar to future avenues for methodological research, there are also plenty of opportunities to extend the empirical analyses presented in this thesis. For one, it would be critical to understand the performance of the presented nowcasting models also for other countries to assess the generalisability of the results. Likewise, the GaR applications have focused only on the US economy. It is well known, however, that risks to economic activity are increasingly shared across borders as our economies have become ever more interconnected (see e.g. Carriero et al. (2020a)). Hence, it would be interesting to extend the analysis to panel or hierarchical settings with the use of appropriate prior extensions to deal with issues of inter-connectedness as well as grouping.

Appendix A

Appendix of chapter 1

A.1 Model Details

A.1.1 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 (2.2.2).

Horseshoe Prior

Starting from model 2.2 and assuming that the states and state variances have already been drawn in steps 1.-3. in 2.2.2 which is further described in (A.1.4) 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:

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

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

$$\begin{aligned}\beta | y^*, \varphi, \vartheta, \sigma &\sim N(A^{-1} X' y^*, \sigma^2 A^{-1}) \\ A &= (X' X + \Lambda_*^{-1}) \\ \Lambda_* &= \vartheta^2 \text{diag}(\varphi_1^2, \dots, \varphi_K^2)\end{aligned}\tag{A.2}$$

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 $\mathcal{O}(T^2 K)$. Suppose the posterior is non-

mal $N_K(\mu, \Sigma)$ with

$$\Sigma = (\phi'\phi + D^{-1})^{-1}, \quad \mu = \Sigma\phi'\alpha, \quad (\text{A.3})$$

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:

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 $\varsigma = u + D\Phi'w$
-

ς represents a direct draw from $N(\mu, \Sigma)$. 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):

$$\sigma^2 | y^*, \beta, \varphi, \vartheta \sim \mathcal{G}^{-1}((T - K)/2, (y^* - X\beta)'(y^* - X\beta)/2 + \beta'\Lambda_*^{-1}\beta/2), \quad (\text{A.4})$$

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

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

$$p(\eta_j | \vartheta, \sigma, \mu_j) \propto \exp(-\frac{\mu_j^2}{2}\eta_j) \frac{1}{1 + \eta_j}. \quad (\text{A.5})$$

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 φ_j yields a direct draw from its posterior distribution and by setting $\eta = 1/\vartheta^2$ and replacing $\mu_j^2 = \sum \beta_j^2/2$ yields an equivalent draw from the conditional posterior of ϑ . 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.

SSVS Prior

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

$$\begin{aligned}
\beta_j | \gamma_j, \delta_j^2 &\sim \gamma_j N(0, \delta_j^2) + (1 - \gamma_j) N(0, c \times \delta_j^2) \\
\delta_j^2 &\sim \mathcal{G}^{-1}(a_1, a_2) \\
\gamma_j &\sim \text{Bern}(\pi_0) \\
\pi_0 &\sim \mathcal{B}(b_1, b_2),
\end{aligned} \tag{A.6}$$

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 $\gamma_j | \beta_j, \delta_j^2, \pi_0, y \sim (1 - \pi) N(\beta_j | 0, c \times \delta_j^2) I_{\gamma_j=0} + \pi_0 N(\beta_j | \beta_j | 0, \delta_j^2) I_{\gamma_j=1}$
 - 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.1.2 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 (2.2.2), the conditional posterior $p(\beta | \Lambda_*, \tau, \sigma^2, y)$ within a normal linear regression can be written as:

$$\bar{\beta}_j = (1 - \kappa_j) \hat{\beta}_j \tag{A.7}$$

where

$$\kappa_j = \frac{1}{1 + T \sigma^{-2} \vartheta^2 s_j^2 \varphi_j^2}, \tag{A.8}$$

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 $\varphi_j\vartheta \rightarrow \infty$, then $\kappa_j \rightarrow 0$ and when $\varphi_j\vartheta \rightarrow 0$, then $\kappa_j \rightarrow 1$. The distribution of the shrinkage coefficient $p(\kappa_j|\Lambda_*, \vartheta, \sigma^2, y)$ is implicitly defined through the priors for the hyperparameters (Λ, ϑ) . 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.

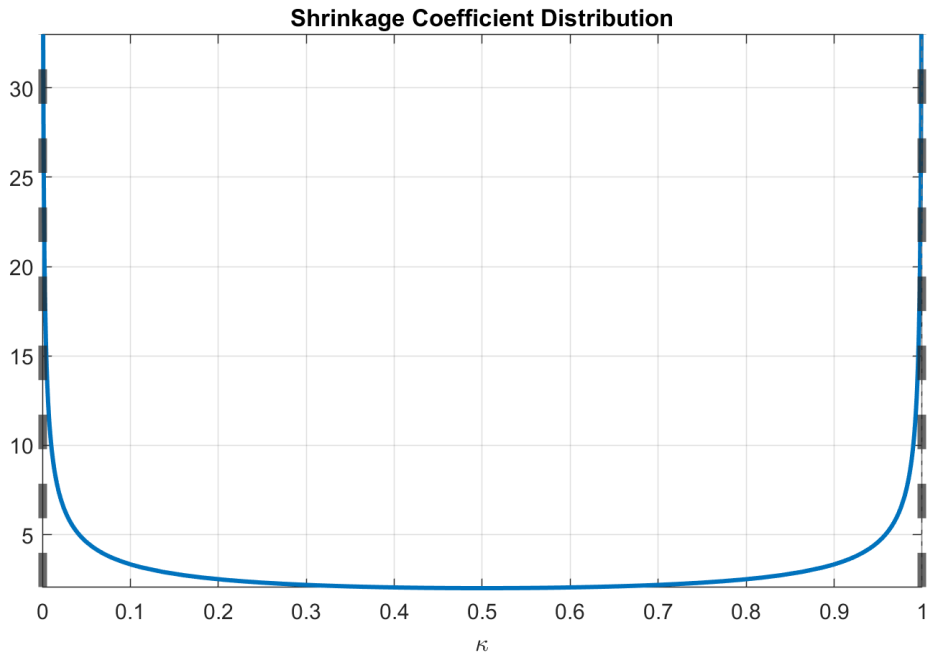


Figure A.1: $\kappa_j|\bullet$ distributions for the horseshoe prior (blue lines) and the spike-and-slab prior (dashed grey lines) with zero point mass and infinite slab scale.

Now, by summing over the shrinkage coefficients in (A.8), 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 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 (A.2).

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

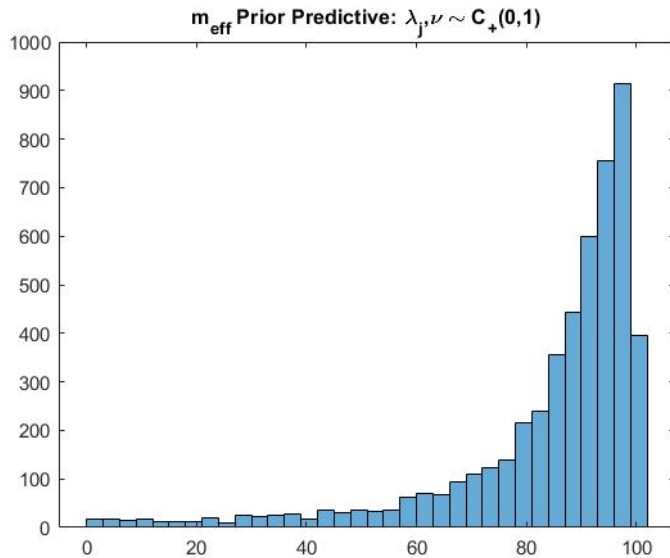


Figure A.2: m_{eff} generated by setting $T = K = 100$ and $\sigma = 1$. Distribution based on 5000 samples according to the prior in (F.2.2).

account for uncertainty in (ϑ, 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.

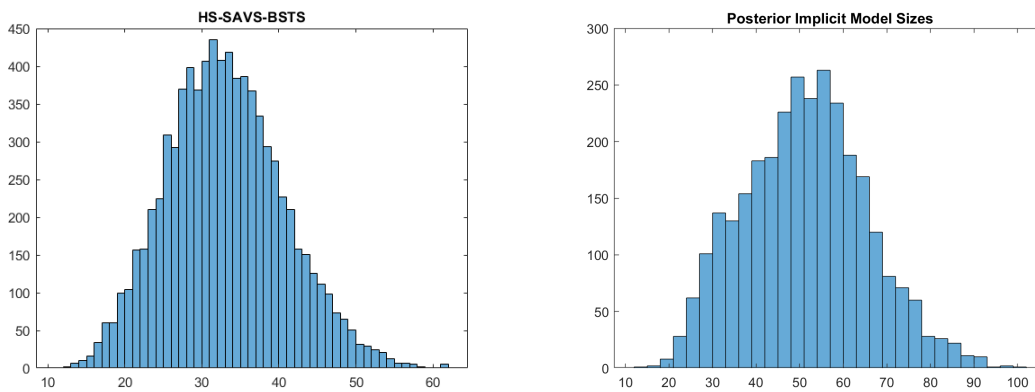


Figure A.3: Posterior model size distribution for the HS-SAVS-BSTS (left) and the posterior of m_{eff} for the HS-BSTS (right)

Nevertheless, we provide in Figure A.3 the posterior effective model size distributions for 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.1.3 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 chapter, but with ARMA components instead of a LLT. The model considered is:

$$y_t = \beta_0 + x_t' \beta + \sum_{j=1}^p \rho_j y_{t-j} + \epsilon_t$$

$$\epsilon_t = u_t + \sum_{m=1}^q \psi_m u_{t-q},$$
(A.9)

where $|\psi_m| < 1$ 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 $\psi_m \sim \mathcal{U}(-1, 1)$. 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.

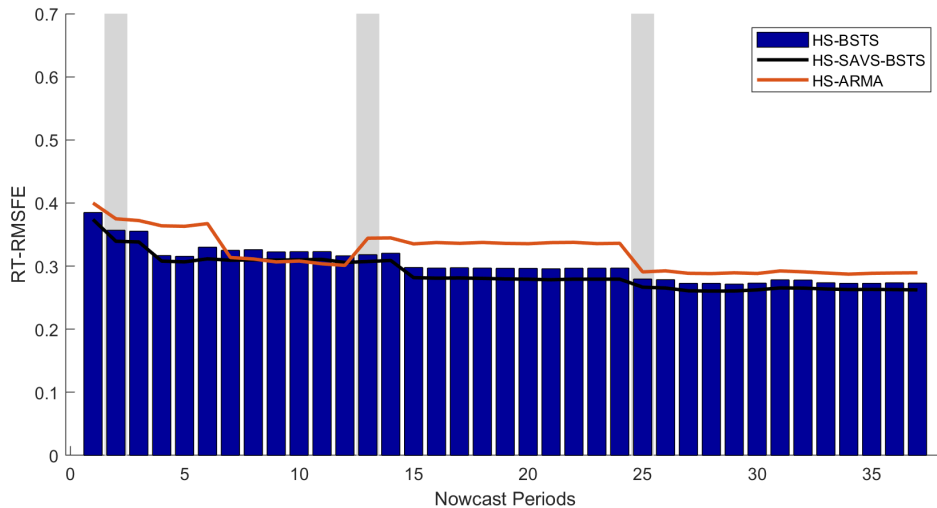


Figure A.4: Real-Time-RMSFE with HS-ARMA.

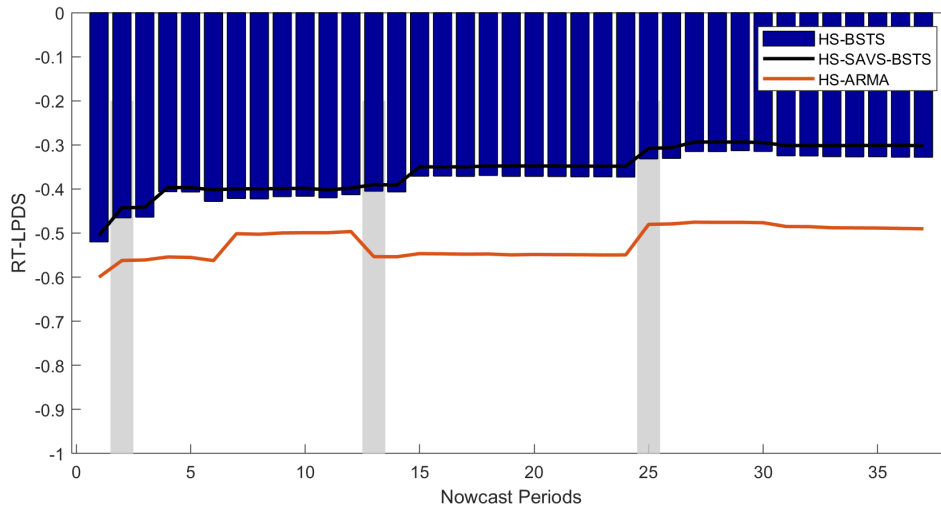


Figure A.5: Real-Time-LPDS with HS-ARMA.

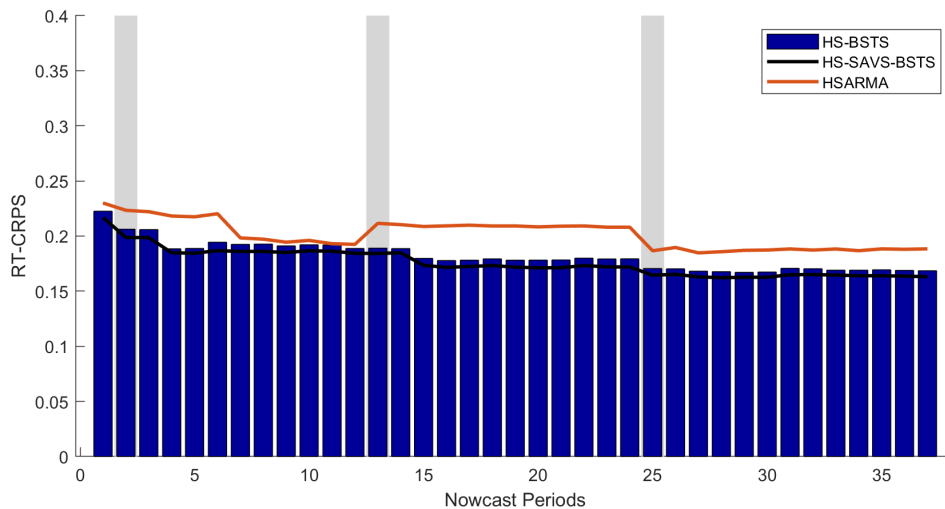


Figure A.6: Real-Time-CRPS with HS-ARMA.

As can be clearly seen from Figures A.4-A.6, 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 (A.7): the LLT component captures smooth transitions in GDP growth, while the 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.

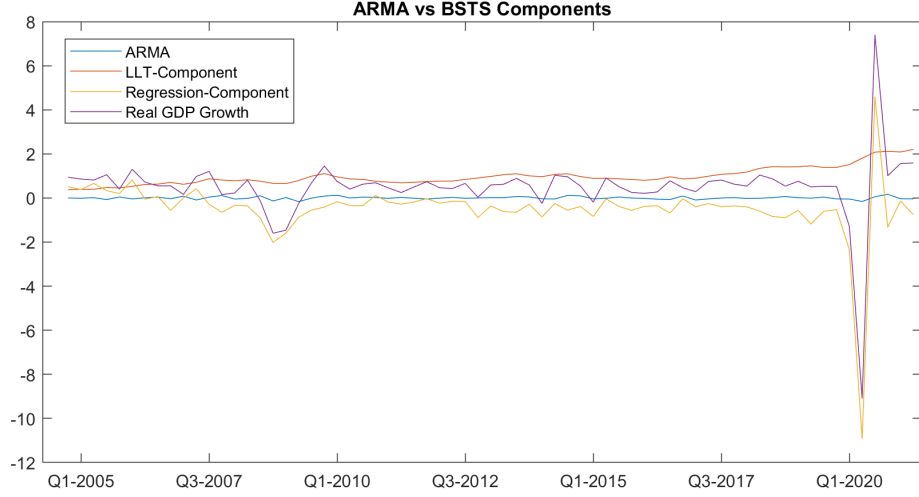


Figure A.7: Trend and regression decomposition for the HS-BSTS model and the best performing HS-ARMA model (ARMA(1,1)).

A.1.4 State Space Estimation and Forecasting

Estimation

Assume analogously to A.1.1 and A.1.1 that all regression parameters have been sampled such that conditionally on β , we estimate $y - X\beta = \hat{y}_t = \tau_0 + \sigma_\tau \tilde{\tau}_t + t\alpha_0 + \sigma_\alpha \sum_{s=1}^t \tilde{\alpha}_s + \epsilon$, $\epsilon_t \sim N(0, \sigma_y)$ 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 fomrulation, 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 (2.3) implies that $\mathbf{H}\tilde{\alpha} = \tilde{\mathbf{u}}^\alpha$, where \mathbf{H} is the first difference matrix 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

$$\tilde{\mathbf{A}} \sim N(\mathbf{0}, (\mathbf{H}^2' \mathbf{H}^2)^{-1}) \quad (\text{A.10})$$

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

$$\hat{y} = \tau_0 \mathbf{1}_T + \alpha_0 \mathbf{1}_{1:T} + \mathbf{X}_\xi \xi + \epsilon, \quad (\text{A.11})$$

where $\mathbf{1}_{1:T}$ is defined as $(1, 2, \dots, T)'$ and $\mathbf{X}_\xi = (\sigma_\tau \mathbf{I}_T, \sigma_\alpha \mathbf{I}_T)$. 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:

$$\xi \sim N(0, \mathbf{P}_\xi^{-1}), \quad (\text{A.12})$$

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

$$p(\xi|\mathbf{y}, \sigma_y^2) \sim N(\bar{\xi}, A_\xi^{-1}) \quad (\text{A.13})$$

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:

$$\begin{aligned} \theta|\mathbf{y}, \sigma_y^2 &\sim N(\bar{\theta}, A_\theta^{-1}) \\ A_\theta &= (V_\theta^{-1} + \frac{1}{\sigma_y^2}X'_\theta X_\theta) \\ \bar{\theta} &= A_\theta^{-1}(V_\theta^{-1}\theta_0 + \frac{1}{\sigma_y^2}X'_\theta \hat{\mathbf{y}}). \end{aligned} \quad (\text{A.14})$$

Forecasting

Taking equation (2.1) 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 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 computational 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.1.5 BSTS-t Model

The model to be estimated is:

$$y_t = \tau_0 + \sigma_\tau \tilde{\tau}_t + t\alpha_0 + \sigma_\alpha \sum_{s=1}^t \tilde{\alpha}_s + x_t' \beta + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma_y^2 \lambda_t), \quad \lambda_t \sim \mathcal{G}^{-1}(\nu/2, \nu/2) \quad (\text{A.15})$$

and

$$\begin{aligned} \tilde{\tau}_t &= \tilde{\tau}_{t-1} + \tilde{u}_t^\tau, \quad \tilde{u}_t^\tau \sim N(0, 1) \\ \tilde{\alpha}_t &= \tilde{\alpha}_{t-1} + \tilde{u}_t^\alpha, \quad \tilde{u}_t^\alpha \sim N(0, 1) \end{aligned} \quad (\text{A.16})$$

Compared to the normal BSTS models,, $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_T)'$ and ν are additional unknown parameters.

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

To derive $p(\beta|\bullet)$, rewrite again $y^* = y - \tau = X\beta + \epsilon$. Then $\epsilon \sim (0_T, \sigma_y^2 \Theta)$, where $\Theta = \text{diag}(\lambda_1, \dots, \lambda_T)$. Defining $\Sigma = \sigma_y^2 \Theta$:

$$\begin{aligned} \beta|\bullet &\sim N(\hat{\beta}_1, D_1) \\ D_1 &= (\Lambda_*^{-1} + X' \Sigma^{-1} X)^{-1} \\ \hat{\beta}_1 &= D_1 X' \Sigma^{-1} y^*, \end{aligned} \quad (\text{A.17})$$

where the diagonal of Λ_* is populated by the shrinkage scales of either the horseshoe prior or the SSVS.¹

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

$$\begin{aligned} \theta|\bullet &\sim N(\hat{\beta}_2, D_2) \\ D_2 &= (V_\theta^{-1} + X_\theta' \Sigma^{-1} X_\theta)^{-1} \\ \hat{\beta}_2 &= D_2 (V_\theta^{-1} \theta_0 + X_\theta' \Sigma^{-1} \hat{y}) \end{aligned} \quad (\text{A.18})$$

To sample from $(\tilde{\tau}, \tilde{\alpha}|\bullet)$, take equation A.11 with $\epsilon_t \sim N(0, \sigma_y^2 \lambda_t)$ and $\lambda_t \sim \mathcal{G}^{-1}(\nu/2, \nu/2)$. Then the posterior with the same steps as above is:

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

$$\begin{aligned}
\xi|\hat{y}, \sigma_y^2 &\sim N(\hat{\beta}_3, D_3) \\
D_3 &= (P_\xi + \frac{1}{\sigma_y^2} X'_\xi \Sigma^{-1} X_\xi)^{-1} \\
\hat{\beta}_3 &= \mathbf{K}_\xi^{-1} (\frac{1}{\sigma_y^2} \mathbf{X}'_\xi \Sigma^{-1} (\mathbf{y} - \tau_0 \mathbf{1}_T - \alpha_0 \mathbf{1}_{1:T})).
\end{aligned} \tag{A.19}$$

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

$$p(\boldsymbol{\lambda}|\bullet) \propto \prod_{t=1}^T \lambda_t^{-\frac{\nu+1}{2}+1} e^{-\frac{1}{2\lambda_t}(\nu + \frac{(y_t - \tau - x'_t \beta)^2}{\sigma_y^2})}. \tag{A.20}$$

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

$$\lambda_t \sim \mathcal{G}^{-1}\left(\frac{\nu+1}{2}, \frac{1}{2} \frac{(y_t - \tau - x'_t \beta)^2}{\sigma_y^2}\right) \tag{A.21}$$

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

$$\begin{aligned}
p(\nu|\bullet) &\propto p(\boldsymbol{\lambda})p(\nu) \\
&\propto \prod_{t=1}^T \frac{(\nu/2)^{\frac{\nu}{2}}}{\Gamma(\nu/2)} \lambda_t^{-(\frac{\nu}{2}+1)} e^{-\frac{\nu}{2\lambda_t}} \\
&= \frac{(\nu/2)^{\frac{T}{2}}}{\Gamma(\nu/2)^T} \left(\prod_{t=1}^T \lambda_t\right)^{-(\frac{\nu}{2}+1)} e^{\frac{\nu}{2} \sum_{t=1}^T \lambda_t^{-1}}
\end{aligned} \tag{A.22}$$

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

$$\alpha(X, Y) = \min\left\{\frac{f(Y)g(X)}{f(X)g(Y)}, 1\right\}, \tag{A.23}$$

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

The sampling algorithm follows the same sequence as in main text of the chapter, however with steps $p(\boldsymbol{\lambda}|\bullet)$ and $p(\nu|\bullet)$ added before sampling $\sigma_y^2|\bullet$ is sampled.

A.2 Graphs

A.2.1 HS-BSTS-t Model Results

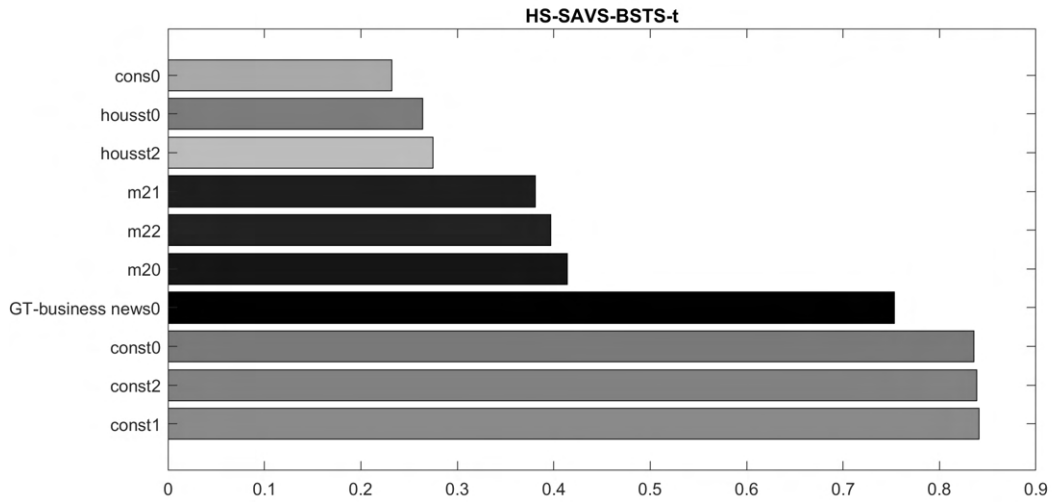


Figure A.8: Posterior inclusion probabilities for the HS-SAVS-BSTS-t model.

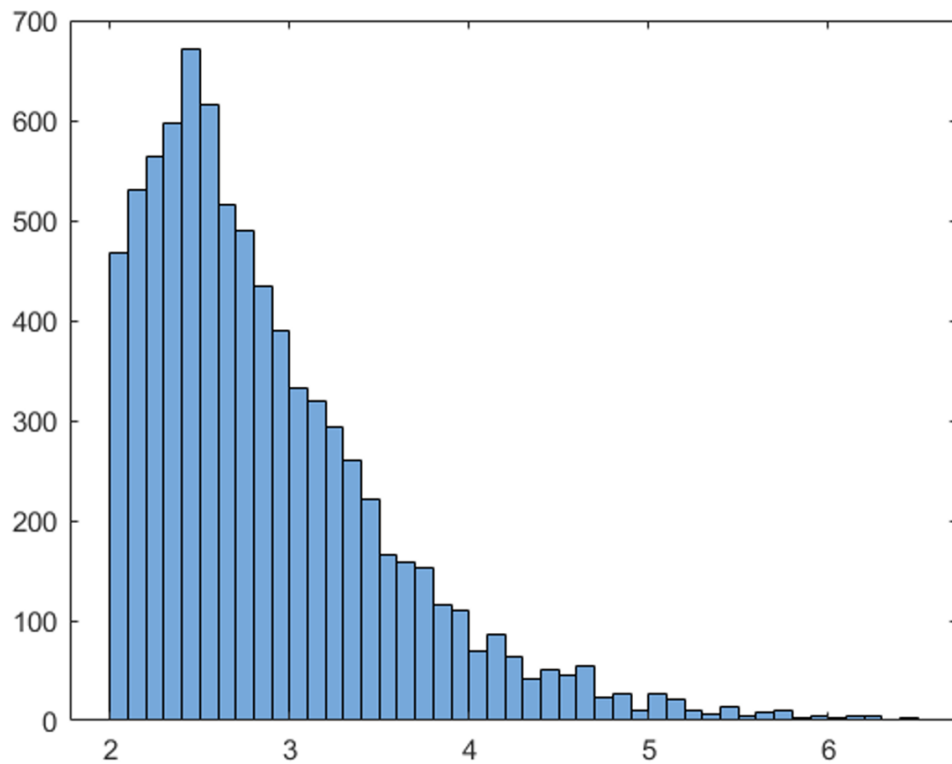


Figure A.9: Posterior of the degree of freedom parameter, ν .

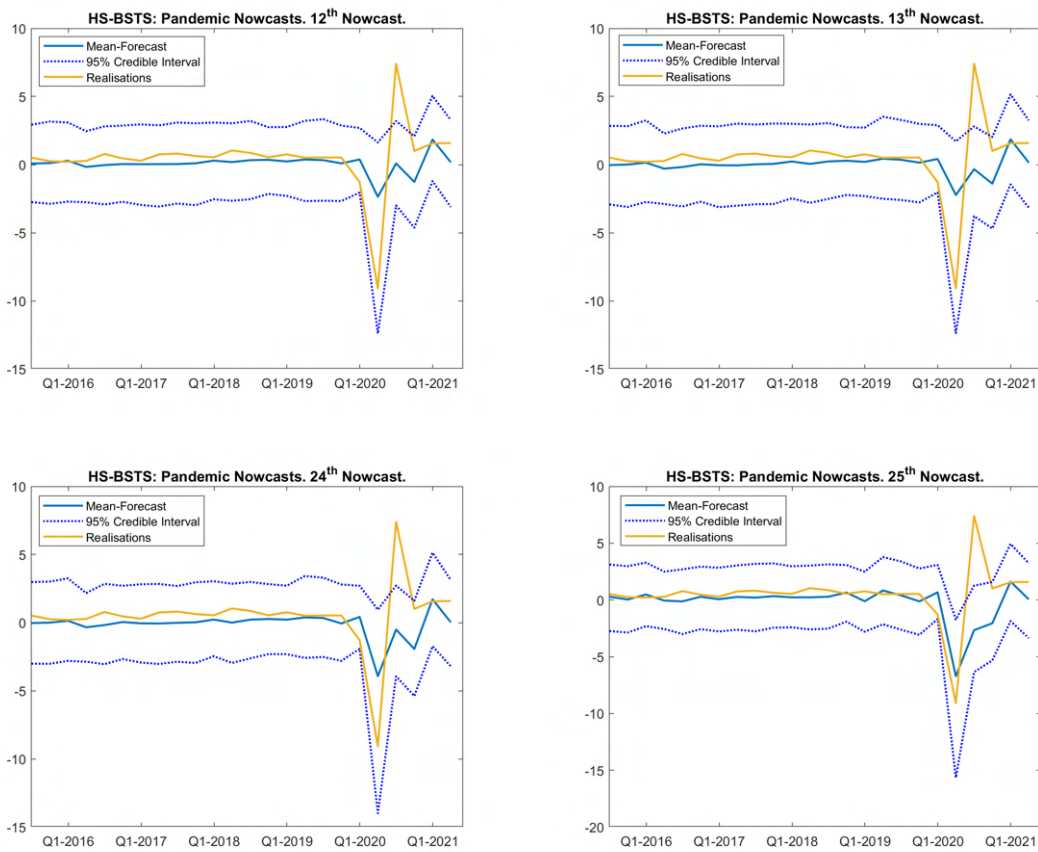


Figure A.10: Predictive distributions for full HS-BSTS-t model for nowcast periods 12,13,24 and 25. Periods 12 and 24 are prior to Google Trends releases.

A.2.2 In-Sample Results

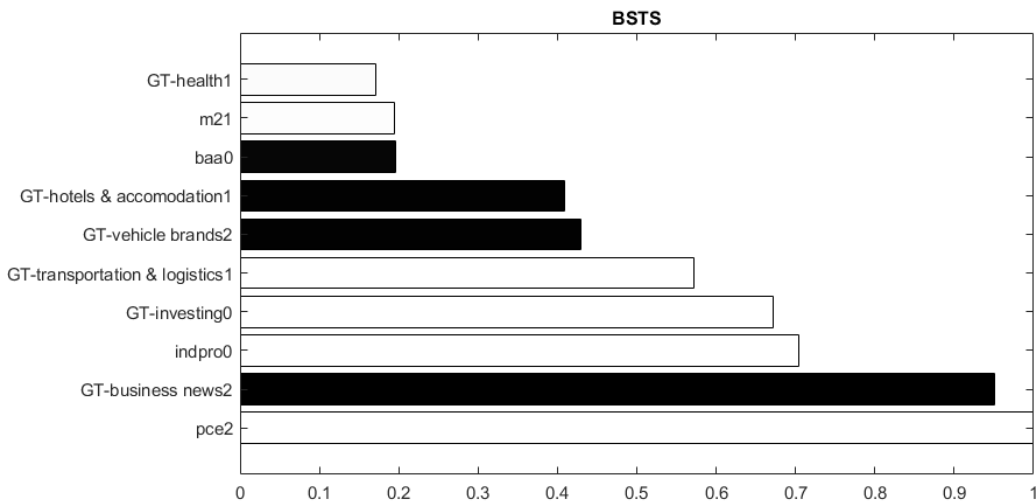


Figure A.11: Posterior inclusion probabilities for the original BSTS model of Scott and Varian (2014).

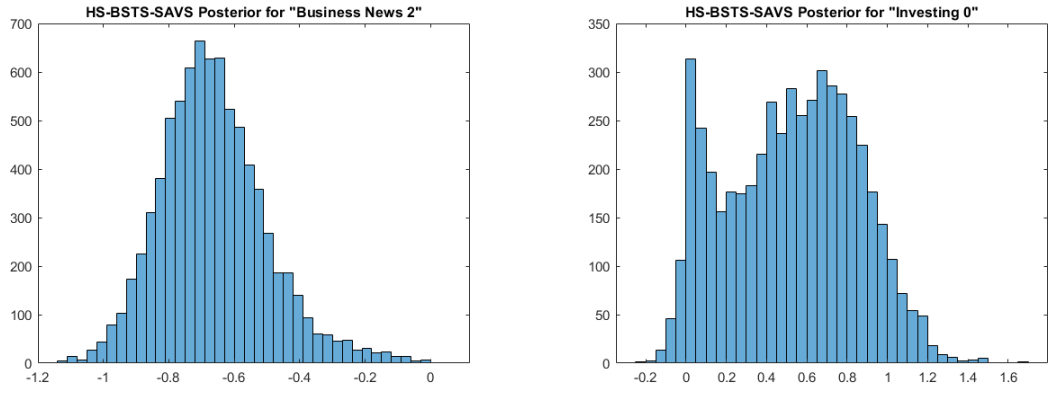


Figure A.12: Posterior of the Google category 'Business news 2' (left) and topic 'Investing 0' for HS-SAVS-BSTS.

A.2.3 Google Topic/Category Plots



Figure A.13: U-MIDAS transformed Google search data plots.

Appendix B

Appendix of Chapter 2

B.1 Model Details

B.1.1 Posteriors

Further Insight into the GIGG Posterior

In order to highlight the effect of different choices on a_k and b_k , we re-write the posterior in (3.6) in terms of its shrinkage coefficient representation:

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

where we define $\mathbf{z}_k = (z'_{k,1}, \dots, z'_{k,T})'$, $\Lambda_{t,h} = \text{diag}(\lambda_1^2 e^{h_1}, \dots, \lambda_T^2 e^{h_T})$ and $\hat{\theta}_{k,j} = (\mathbf{z}_k^{(m)'} \Lambda_{t,h}^{-1} \mathbf{z}_k^{(m)})^{-1} \mathbf{z}_k^{(m)'} \Lambda_{t,h}^{-1} \tilde{\mathbf{y}}$ can be viewed as a conditional maximum likelihood estimate for $\theta_{k,j}$. We suppress the indication of monthly frequency (m) for convenience here. Note that for this representation, we have assumed that the maximum likelihood estimate for group k exists and that the $p_k + 1$ lags within \mathbf{z}_k have been orthogonalised. Since the group-size is likely to be much smaller than the sample size, and given that we will group-orthogonalise the the data anyway (see section 3.2.3), these only represent very mild assumptions. Under these assumptions, it is easy to verify that the shrinkage coefficient $\kappa_{k,j|\bullet} = \frac{1}{1 + \tilde{s}_{k,j}^2 \vartheta^2 \gamma_k^2 \varphi_{k,j}^2}$ is bounded between 0 and 1 and thus dictates how far away the prior shrinks the coefficients from the maximum likelihood solution. It is easy to see that $\vartheta^2 \gamma_k^2 \varphi_{k,j}^2 \rightarrow \infty$, $\bar{\theta}_{k,j} \rightarrow \hat{\theta}_{k,j}$. The distribution $\pi(\kappa_{k,j})$ which is implicitly defined via the priors on γ_k^2 and $\varphi_{k,j}^2$ determine the a-priori shrinkage behaviour we can expect. By assuming $\gamma_k^2 \varphi_{k,j}^2 \sim \beta'(a_k, b_k)$, the joint distribution for $\boldsymbol{\kappa}_k$ can be factored as:

$$\pi(\kappa_k|\bullet) = \frac{\Gamma(a_k + (p_k + 1)b_k)}{\Gamma(a_k)\Gamma(b_k)^{k+1}} \prod_{j=1}^{p_k} \tilde{s}_{k,j}^{b_k} \left(1 + \sum_{j=1}^{p_k} \tilde{s}_{k,j} \frac{\kappa_{k,j}}{(1 - \kappa_{k,j})}\right)^{-(a_k + (1+p_k)b_k)} \times \left(\prod_{j=1}^{p_k} \kappa_{k,j}^{b_k-1} (1 - \kappa_{k,j})^{-(b_k+1)}\right), \quad (\text{B.2})$$

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

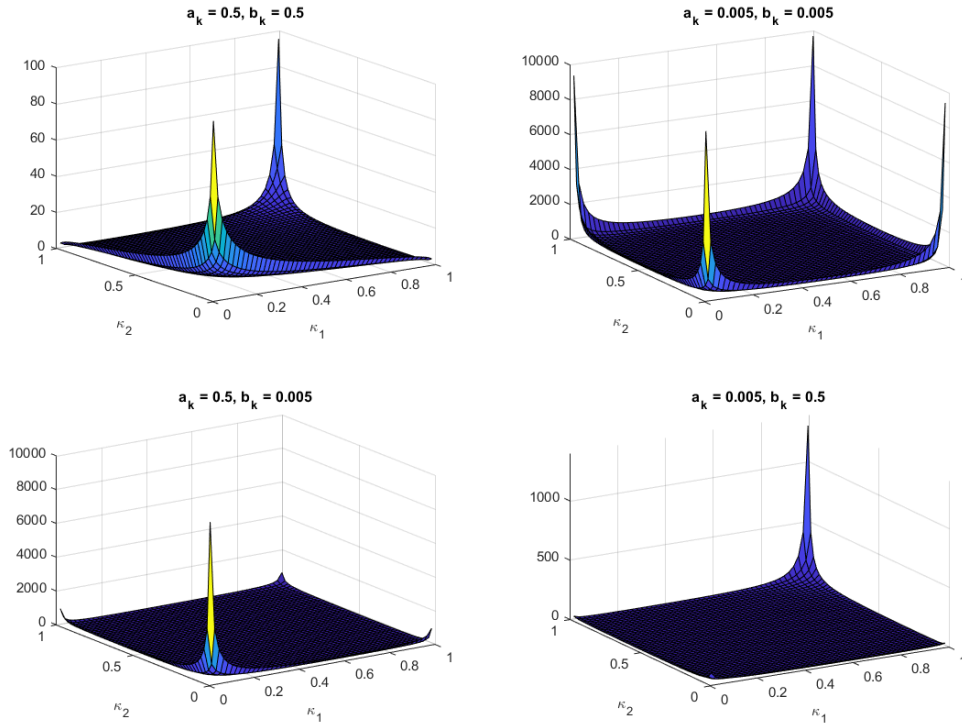


Figure B.1: Bi-variate shrinkage coefficient plots for various hyper-parameter values. a_k controls group-level sparsity while b_k controls the degree of correlation with the overall sparsity level.

As expected, when b_k is relatively small compared to a_k , then the sparsity level enforced by a_k dominates: lower left hand panel showcases a situation in which relatively little shrinkage is exerted because a_k is relatively large compared to b_k , while the upper right hand panel's joint distribution is characterised by an independently distributed, very extreme horseshoe behaviour (the U shape is much narrower than that implied by the standard horseshoe). When b_k instead is relatively large compared to a_k , then the shrinkage behaviour will tend to be symmetric. The exact group-horseshoe case ($a_k = b_k$), for example, resembles a joint U-shape. Hence, for the exact group-horseshoe, a-priori either the entire group will be shrunk to zero, or

all coefficients are left relatively un-perturbed. This exposition will help understand which behaviour to expect given the choice of the hyper-parameters for the nowcast application in the empirical application.

Posteriors of Hyper-parameters

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

$$(\vartheta | \mathbf{y}, \bullet) \sim \mathcal{G}^{-1}\left(\frac{\sum_{k=1}^K (p_k + 1) + 1}{2}, \theta' \Lambda_p^{-1} \theta / 2 + \frac{1}{\xi_\beta}\right) \quad (\text{B.3})$$

$$(\gamma_k^{-2} | \mathbf{y}, \bullet) \sim \mathcal{GiG}\left(\frac{p_k + 1}{2} - a_k, \frac{1}{\vartheta^2} \sum_{j=1}^{p_k+1} \frac{\theta_{kj}^2}{\varphi_{kj}^2}\right) \quad (\text{B.4})$$

$$(\varphi_{kj}^2 | \mathbf{y}, \bullet) \sim \mathcal{G}^{-1}\left(b_k + \frac{1}{2}, 1 + \frac{\theta_{kj}^2}{2\vartheta^2 \gamma_k^2}\right) \quad (\text{B.5})$$

$$(\xi_\beta | \mathbf{y}, \bullet) \sim \mathcal{G}^{-1}\left(1, \frac{1}{\vartheta}\right), \quad (\text{B.6})$$

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

Posteriors of the State Space

In this section, we will detail the conditional posteriors of each of the remaining parameters of model which will be used to construct the Gibbs sampler in. For convenience, we reproduce the main Trend-SV-t model of section 3.2 here again:

$$\begin{aligned} y_t &= \tau_t + \theta' Z_{t-h}^{(m)} + \sqrt{\lambda_t} e^{\frac{1}{2}(h_0 + w_h \tilde{h}_t)} \tilde{\epsilon}_t^y, \\ \tilde{\epsilon}_t^y &\sim N(0, 1), \quad \lambda_t \sim \mathcal{G}^{-1}(\nu/2, \nu/2) \end{aligned} \quad (\text{B.7})$$

$$\begin{aligned} \tau_t &= \tau_{t-1} + e^{\frac{1}{2}(g_0 + w_g \tilde{g}_t)}, \quad \tilde{\epsilon}_t^g \sim N(0, 1) \\ \tilde{h}_t &= \tilde{h}_{t-1} + \tilde{\epsilon}_t^h, \quad \tilde{\epsilon}_t^h \sim N(0, 1), \quad \tilde{h}_0 = 0 \\ \tilde{g}_t &= \tilde{g}_{t-1} + \tilde{\epsilon}_t^g, \quad \tilde{\epsilon}_t^g \sim N(0, 1), \quad \tilde{g}_0 = 0 \end{aligned} \quad (\text{B.8})$$

$$\begin{aligned}
w_g &\sim N(0, V_{w_g}), & w_h &\sim N(0, V_{w_h}) \\
h_0 &\sim N(a_{0,h}, b_{0,h}), & g_0 &\sim N(a_{0,g}, b_{0,g}) \\
\tau_0 &\sim N(a_{0,\tau}, b_{0,\tau})
\end{aligned} \tag{B.9}$$

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

$$\begin{aligned}
h_t &= h_0 + w_h \tilde{h}_t \\
g_t &= g_0 + w_g \tilde{g}_t
\end{aligned} \tag{B.10}$$

Posterior of τ

To derive the joint posterior of τ , we make use of the methods proposed by Chan and Jeliazkov (2009); McCausland et al. (2011) as they enable sampling of all states (τ_1, \dots, τ_T) simultaneously. Compared to more traditional forward-sampling-backwards-smoothing algorithms of Carter and Kohn (1994); Durbin and Koopman (2002) which sample states one time-step at a time, this represents an improvement in statistical efficiency as well as computational efficiency. The computational efficiency comes from the special band-matrix form of the resultant state posterior which allows for the use of very efficient sparse matrix operations (Chan and Jeliazkov, 2009).

Since we only need the conditional posterior $\pi(\tau | \tau_0, w_\tau, \bullet)$, we proceed similar to the exposition in 3.6, by defining the relevant conditional likelihood based on $\mathbf{y}^* = \mathbf{y} - \theta \mathbf{Z}^{(m)}$, hence the observations of the target, accounted for the cyclical component. It follows from B.7 that then:

$$\mathbf{y} | \tau, \bullet \sim N(\tau, \Lambda_h \Lambda_t^{1/2}). \tag{B.11}$$

To derive the implicit prior on τ , start by vectorising the state process τ in B.8:

$$H\tau = \tilde{\alpha}^\tau + \eta^\tau, \tag{B.12}$$

where $\tau = (\tau_1, \dots, \tau_T)'$, $\tilde{\alpha}^\tau = (\tau_0, 0, \dots, 0)'$, $\eta^\tau \sim N(0, \Lambda_k)$, $\Lambda_k = \text{diag}(e^{g_1}, \dots, e^{g_T})$ and H is the first difference matrix. From B.12, one can write the joint prior of τ as:

$$\boldsymbol{\tau}|\tau_0, \bullet \sim N(\tau \mathbf{1}_T, (H' \Lambda_k^{-1} H)^{-1}), \quad (\text{B.13})$$

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

$$\boldsymbol{\tau}|\mathbf{y}, \tau_0, \bullet \sim N(\hat{\boldsymbol{\tau}}, K_\tau^{-1}), \quad (\text{B.14})$$

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

To sample τ_0 , recall that it only appears in $\tau_1 = \tau_0 + e^{\frac{1}{2}(g_0 + w_g * \tilde{g}_1)} \tilde{\epsilon}_1^\tau$. Hence, assuming with the independent prior in B.9, the posterior is:

$$\tau_0|\mathbf{y}, \bullet \sim N(\hat{\tau}_0, K_{\tau_0}^{-1}), \quad (\text{B.15})$$

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

Posterior of \mathbf{h}

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

$$\mathbf{y}^+ = h_0 \mathbf{1}_T + w_h \tilde{\mathbf{h}} + \tilde{\epsilon}^{y^+}. \quad (\text{B.16})$$

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

$$\tilde{\mathbf{h}}|\mathbf{y}, \bullet \sim N(\hat{\tilde{\mathbf{h}}}, K_{\tilde{\mathbf{h}}}^{-1}), \quad (\text{B.17})$$

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

The remaining conditional posteriors associated with \mathbf{h} are those of h_0 and w_h . Notice from B.16, conditional on $\tilde{\mathbf{h}}$, d_s and Ω_s , the joint posterior of h_0 and w_h can be found using simple regression results. Define $X_h = (\mathbf{1}_T', \tilde{\mathbf{h}})'$, $\zeta_h = (h_0, w_h)$, $a_{0,\zeta_h} = (a_{0,h}, 0)$, $b_{0,\zeta_h} = \text{diag}(b_{0,h}, V_{w_h})$, then $\zeta_h \sim N(a_{0,\zeta_h}, b_{0,\zeta_h})$. The joint posterior is:

$$\zeta_h | \mathbf{y}, \bullet \sim N(\hat{\zeta}_h, K_{\zeta_h}^{-1}), \quad (\text{B.18})$$

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

Posterior of \mathbf{g}

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

Posterior of λ and ν

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

$$p(\boldsymbol{\lambda} | \bullet) \propto \prod_{t=1}^T \lambda_t^{-\frac{\nu+1}{2}+1} e^{-\frac{1}{2\lambda_t}(\nu + \frac{(y_t - \tau_t - \theta' Z_{t-h}^{(m)})^2}{\exp(h_t)})}. \quad (\text{B.19})$$

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

$$\lambda_t \sim \mathcal{G}^{-1}\left(\frac{\nu+1}{2}, \frac{1}{2} \frac{(y_t - \tau_t - \theta' Z_{t-h}^{(m)})^2}{\exp(h_t)}\right) \quad (\text{B.20})$$

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

$$\begin{aligned} p(\nu | \bullet) &\propto p(\boldsymbol{\lambda})p(\nu) \\ &\propto \prod_{t=1}^T \frac{(\nu/2)^{\frac{\nu}{2}}}{\Gamma(\nu/2)} \lambda_t^{-(\frac{\nu}{2}+1)} e^{-\frac{\nu}{2\lambda_t}} \\ &= \frac{(\nu/2)^{\frac{T}{2}}}{\Gamma(\nu/2)^T} \left(\prod_{t=1}^T \lambda_t\right)^{-(\frac{\nu}{2}+1)} e^{\frac{\nu}{2} \sum_{t=1}^T \lambda_t^{-1}} \end{aligned} \quad (\text{B.21})$$

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

$$\alpha(X, Y) = \min \left\{ \frac{f(Y)g(X)}{f(X)g(Y)}, 1 \right\}, \quad (\text{B.22})$$

where $g(\cdot)$ is the proposal density. In order for the Metropolis-Hastings sampler to

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

B.1.2 Sampling Algorithm

In order to estimate the Trend-SV-t-BMIDAS model with GIGG prior, we make use of a Metropolis-within-Gibbs sampler. With the posterior distributions described in sections B.1.1, B.1.1, 3.6, in hand, we sequentially sample from the following posterior distributions:

1. Sample $\theta|\bullet \sim p(\theta|\mathbf{y}, \bullet)$
2. Sample hyper-parameters $\vartheta, \gamma_k^2, \varphi_{kj}^2, \xi_\beta$ in one block
 - (a) $\vartheta^2 \sim p(\vartheta^2|\mathbf{y}, \bullet)$
 - (b) $\gamma_k^2 \sim 1/p(\gamma_k^{-2}|\mathbf{y}, \bullet)$
 - (c) $\varphi_{kj}^2 \sim p(\varphi_{kj}^2|\mathbf{y}, \bullet)$
 - (d) $\xi_\beta \sim p(\xi_\beta|\mathbf{y}, \bullet)$
3. sample $\tilde{\tau} \sim p(\tilde{\tau}|\mathbf{y}, \bullet)$ 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 $\xi_\beta \sim p(\xi_\beta|\mathbf{y}, \bullet)$ with a Metropolis step as described after equation B.21

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

B.1.3 Group-Selection Algorithm

This section gives further details on the derivation of group-variable selection algorithm in 3.8. For convenience, we replicate the objective function here again, omitting notation involving m for clarity:

$$\mathcal{L}(\tilde{\mathbf{Y}}, \psi) = \frac{1}{2} \|\mathbf{Z}\psi - \tilde{\mathbf{Y}}\|_2^2 + \sum_{k=1}^K \zeta_k \|\psi_k\|_2, \quad (\text{B.23})$$

For simplicity, assume that the predictions $\tilde{\mathbf{Y}}$ can be decomposed as $\tilde{\mathbf{Y}} = \mathbf{Z}\theta + \tilde{\boldsymbol{\epsilon}}^{\mathbf{y}}$, $\tilde{\boldsymbol{\epsilon}}^{\mathbf{y}} \sim N(0, \Sigma)$. Intuitively, objective function 3.8 pushes those ψ_k to zero which have little influence on the predictions of our model, $\tilde{\mathbf{Y}}$.

We take the expectation with respect to 1) the expected risk and 2) $\theta|\mathbf{y}$ to account for all sources of uncertainty of the model¹:

$$\begin{aligned} \mathcal{L}(\tilde{\mathbf{Y}}, \psi) &= E_{\tilde{\mathbf{Y}}|\bullet}[\tilde{\mathbf{Y}}, \psi|\bullet] \\ &= \sum_{k=1}^K \zeta_k \|\psi_k\|_2 + \frac{1}{2} \|\mathbf{Z}\psi - \mathbf{Z}\theta\|_2^2 + \frac{1}{2} \text{tr}(\Sigma) \end{aligned} \quad (\text{B.24})$$

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

$$\begin{aligned} \mathcal{L}(\theta, \psi) &= E_{\theta|\mathbf{y}}[\mathcal{L}(\tilde{\mathbf{Y}}, \psi)] \\ &= \frac{1}{2} \|\mathbf{Z}\psi - \mathbf{Z}\bar{\theta}\|_2^2 + \frac{1}{2} \text{tr}(\mathbf{Z}'\mathbf{Z}\Sigma_\theta) + \sum_{k=1}^K \zeta_k \|\psi_k\|_2, \end{aligned} \quad (\text{B.25})$$

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

$$\mathcal{L}(\theta, \psi) = \frac{1}{2} \|\mathbf{Z}\psi - \mathbf{Z}\bar{\theta}\|_2^2 + \sum_{k=1}^K \zeta_k \|\psi_k\|_2. \quad (\text{B.26})$$

Notice, that we follow Chakraborty et al. (2020); Huber et al. (2019) by solving B.25 on a Gibbs iteration bases (instead over the average of the posterior). Traditional solution methods such as the coordinate descent (Friedman et al., 2010) iteratively solve the sub-gradients L times for each group k until convergence:

$$\psi_k^l = (\|r_k^{(l-1)}\|_2^2 - \zeta_k)_+ \frac{r_k^{(l-1)}}{\|r_k^{(l-1)}\|_2}, \quad l = 1, \dots, L, \quad (\text{B.27})$$

where r_k is the partial residual based on the previous iteration, $r^{(l)} = \mathbf{Z}'_k(\mathbf{y} - \mathbf{Z}_{-k}\psi_{-k}^{(l-1)})$ and $-k$ refers to all but the k^{th} group. Orthonormalising \mathbf{Z}_k via its SVD decomposition (which can conveniently be adapted as in (Breheny and Huang, 2015) when $p_k + 1 \gg T$), and stopping the coordinate descent after one iteration as per Ray and Bhattacharya (2018); Chakraborty et al. (2020) such that $\|r_k\|_2 = \|\theta_k\|_2^2$ results immediately in 3.9.

¹ Since ψ is independent of Σ , the integration of posterior uncertainty of Σ results in a constant and thus does not further influence the optimisation problem

B.1.4 Almon-Lags Restricted MIDAS

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

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

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

Almon lag polynomial restrictions, proposed for economic prediction models by Almon (1965), confine the regression problem to $\beta_{k,l} = \theta_{k,i} l^i$ for lags $l = (1, \dots, L_k)$ and parameters of the polynomial $i = (0, \dots, p_k)$. For a third degree polynomial and $L_k + 1 = 6$, $\beta_k = (\theta_0, \sum_{i=0}^{p_k} \theta_i, \sum_{i=0}^{p_k} \theta_i 2^i, \sum_{i=0}^{p_k} \theta_i 3^i, \sum_{i=0}^{p_k} \theta_i 4^i, \sum_{i=0}^{p_k} \theta_i 5^i)'$. With this, the free parameters are reduced from L_k to p_k for each k , which when $p_k \ll L_k$ induces parsimony. This, however, comes at the cost of bias, that may be non-vanishing with increasing sample size Andreou et al. (2010). These restrictions imply the $(i + 1)^{\text{th}}$ row of Ξ_k equal to $[0^i, \dots, L_k^i]$.

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

² Notice that we suppress the notation indicating that the covariate set is in monthly frequency for readability

B.2 Supplementary Material on the Empirical Application

B.2.1 Further Detail on Data

Table B.1: Further Details on Macro Data

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

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

B.2.2 In-Sample Description

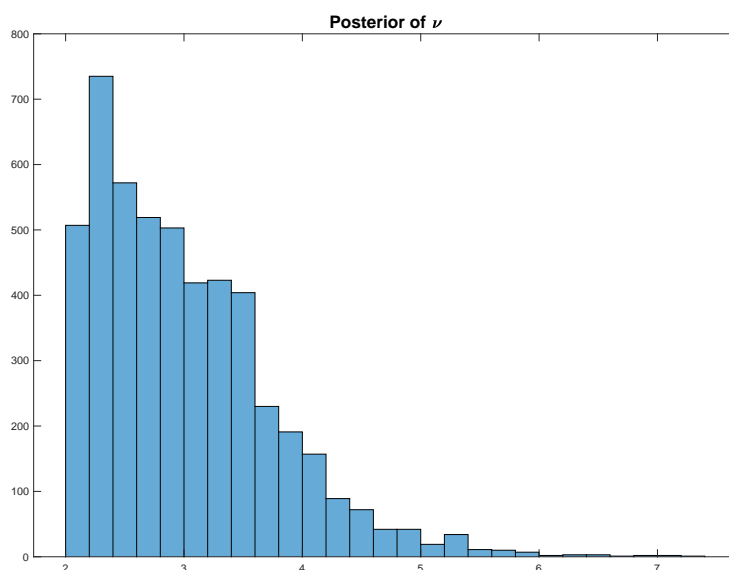


Figure B.2: Posterior Degrees of Freedom of the t-distribution, based on the entire in-sample and last nowcast period.

Figure B.2 has large mass over small posterior degrees of freedom for the t-distribution assumed for the errors of the observation equation of model [reference of equation 1]. The smaller the degrees of freedom, the more leptokurtic the tails of the observation equation's error distribution. Figure B.2 shows strong identification of fat-tails.

B.2.3 Nowcast Performance GIGG Models

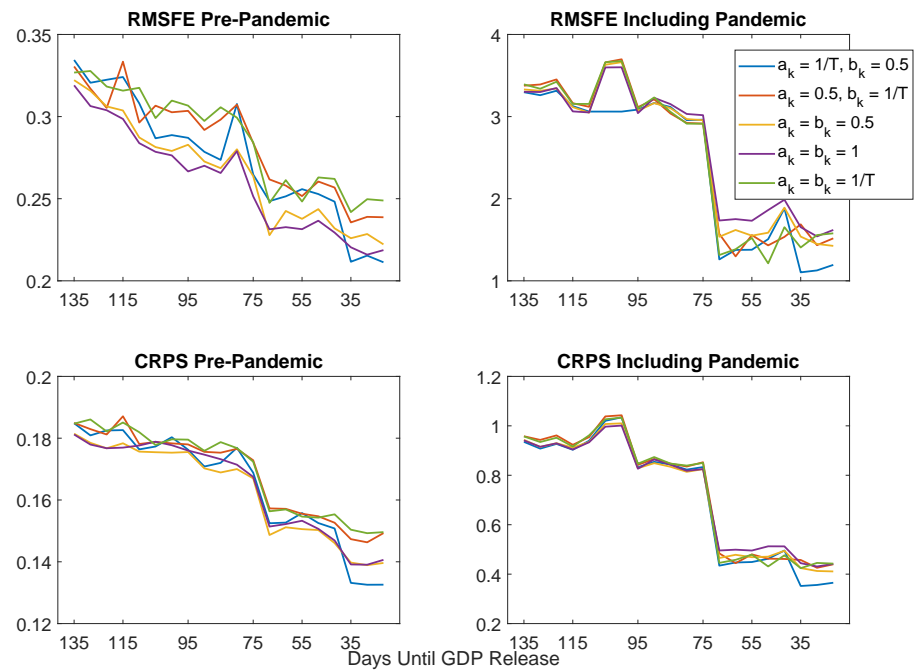


Figure B.3: RSMFE and CRPS pre-pandemic and including pandemic for the Trend-SV-t-GIGG model with different hyper-parameter combinations. $1/T$ is adjusted to the in-sample length at each quarter the nowcasts are conducted. The GIGG prior implemented in the empirical application has $a_k = 1/T, b_k = 0.5$.

B.2.4 Inclusion Probabilities



Figure B.4: Combination weights of the Combination model of high-frequency variables. Weights pre-pandemic don't change much quarter to quarter, so are averages across quarters until Q4-2019. Each subsequent panel shows the weights for each quarter separately. Vertical axis in each panel represents a nowcast period, according to the pseudo-publication table.

B.2.5 Further Simulation Results

Table B.2: Monte Carlo Simulations: DGP 2

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

The table shows in bold the level of each of the evaluation metrics, while those of the GAL-SS and GIGG are in relative terms to the corresponding entry of the horseshoe model. $\bar{\sigma}$ is the Monte Carlo error of the DGP.

Appendix C

Appendix of Chapter 3

C.1 Derivations and Implementation Details

C.1.1 Derivation Algorithm 1

We now give further details on the derivation of Algorithm 1. The goal of the algorithm is to circumvent having to compute large $K \times K$ matrices by redefining auxiliary variables which under certain linear combination result in draws of the desired distribution $N(\bar{\beta}_p, \bar{\Lambda}_*^{-1})$. As above, by the Sherman-Morrison-Woodbury theorem (see e.g. Hager 1989), Θ and μ can be expanded as:

$$\Theta = (\Phi'\Phi + D^{-1})^{-1} = D - D\Phi'(\Phi D\Phi' + I_T)^{-1}\Phi D$$

$$\mu = D\Phi'(\Phi D\Phi' + I_T)^{-1}\alpha,$$

Where the second equality follows from applying the push-through identity after some re-arranging of terms. This expansion per-se won't help in sampling from $N(0, \Theta)$. Letting v and u being defined as in chapter 4, $w = (v', u')' \in \mathbb{R}^{T+K}$ follows a multivariate normal distribution centred on 0 with covariance

$$\Omega = \begin{pmatrix} P & S \\ S' & D \end{pmatrix}$$

where it is easily verified that $P = (\Phi D\Phi' + I_T)$ and S' can be derived as:

$$\begin{aligned} \text{Cov}(v, u) &= \text{Cov}(D^{1/2}\epsilon, \Phi u) \\ &= E(D^{1/2}\epsilon u' X' U^{1/2}) \\ &= E(D^{1/2}\epsilon \epsilon' D^{1/2} X' U^{1/2}) \\ &= D X' U^{1/2} \\ &= D\Phi' \end{aligned}$$

where $\epsilon \sim N(0, I_K)$. The second equality follows from the fact that both quantities are independently centred on 0. Rewriting Ω into its LDU decomposition (see e.g. Hamilton, 1994) as:

$$\begin{pmatrix} P & S \\ S' & D \end{pmatrix} = \underbrace{\begin{pmatrix} I_T & 0 \\ S'P^{-1} & I_K \end{pmatrix}}_L \underbrace{\begin{pmatrix} P & 0 \\ 0 & D - S'P^{-1}S \end{pmatrix}}_\Gamma \underbrace{\begin{pmatrix} I_T & P^{-1}S \\ 0 & I_K \end{pmatrix}}_{L'}$$

Where the lower $K \times K$ block in Γ is equal to Θ using the Sherman-Morrison-Woodbury identity. To retrieve the lower part, we isolate Γ which is easily obtained because L is lower triangular and thus the inverse is readily available as:

$$L^{-1} = \begin{pmatrix} I_T & 0 \\ -S'P^{-1} & I_K \end{pmatrix}$$

Since w has already been sampled from $N(0, \Omega)$ in steps 2 and 3 of the algorithm, the transformation $w_* = L^{-1}w$ is distributed $N(0, \Gamma)$. Collecting the lower block of w_* yields a sample from $N(0, \Theta)$. Finally, by adding the α term in step 3 of the algorithm, we center the combined distribution on μ which completes the algorithm.

C.1.2 Slice Sampling

Slice sampling generates pseudo-random numbers from any distribution function $f(y)$ by sampling uniformly from horizontal slices through the PDF. Advantages of the algorithm include its simplicity, that it involves no rejections, and that it requires no external parameters to be set. Define $\eta_j = 1/\varphi_j^2$ and $\mu_j = \beta_j/\vartheta$. The conditional posterior distribution of η_j , given all other parameters is given by

$$p(\eta_j|\vartheta, \Theta, \mu_j, \theta, \tau, X, Y, Z) \propto \exp\left\{-\frac{\mu_j^2}{2}\eta_j\right\} \frac{1}{1+\eta_j}$$

Slice sampling can now be implemented to draw from (14):

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

Taking the inverse square root of the sample of 2., one receives back the estimate for φ_j . By replacing $\eta = 1/\vartheta$ and μ_j^2 by $\sum_{j=1}^K (\beta_j/\varphi_j)^2/2$, ϑ can be sampled in a similar manner.

C.2 Supplementary Material on Empirical Application

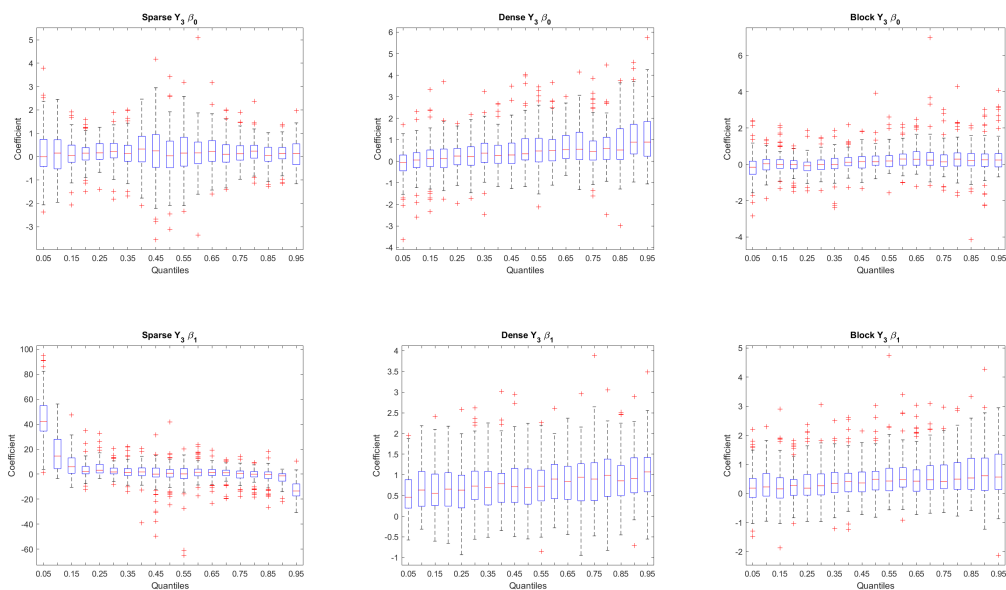


Figure C.1: LBQR's β_0 and β_1 profiles for y_3 across quantiles for the different sparsity settings

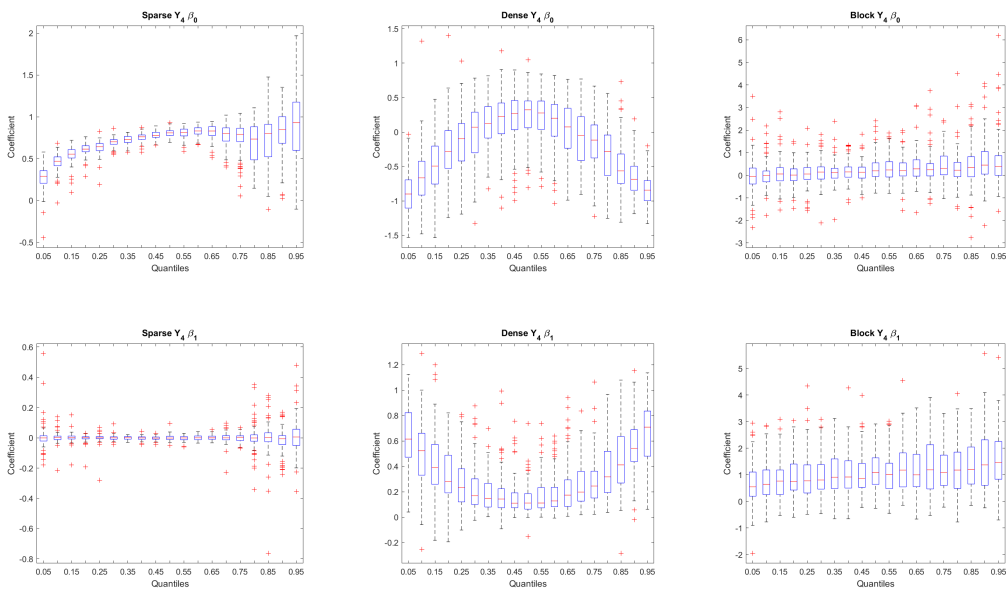


Figure C.2: LBQR's β_0 and β_1 profiles for y_4 across quantiles for the different sparsity settings

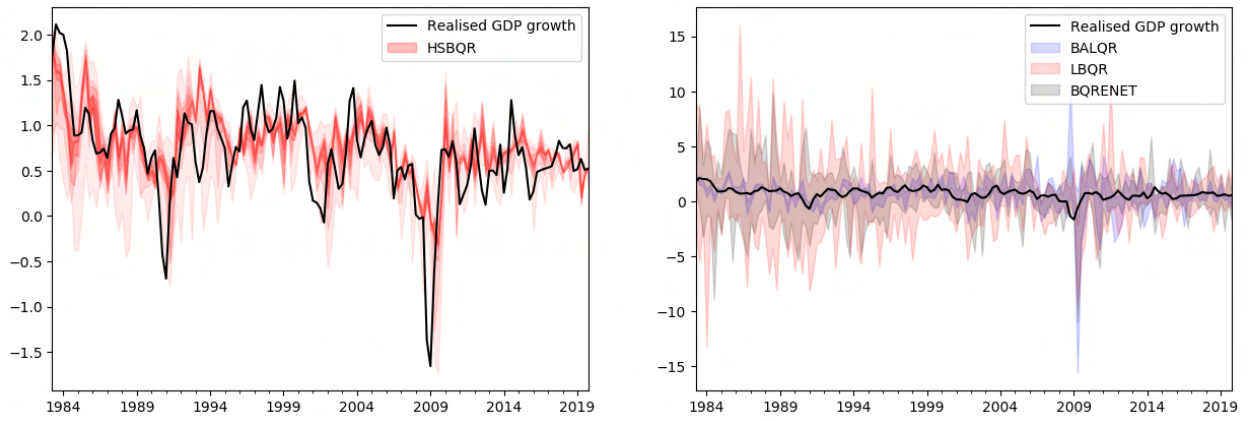


Figure C.3: Two-step-ahead forecast distributions for the L1QR, BQR, BALQR and HS-BQR. Shaded areas correspond to plots of all 19 quantiles.

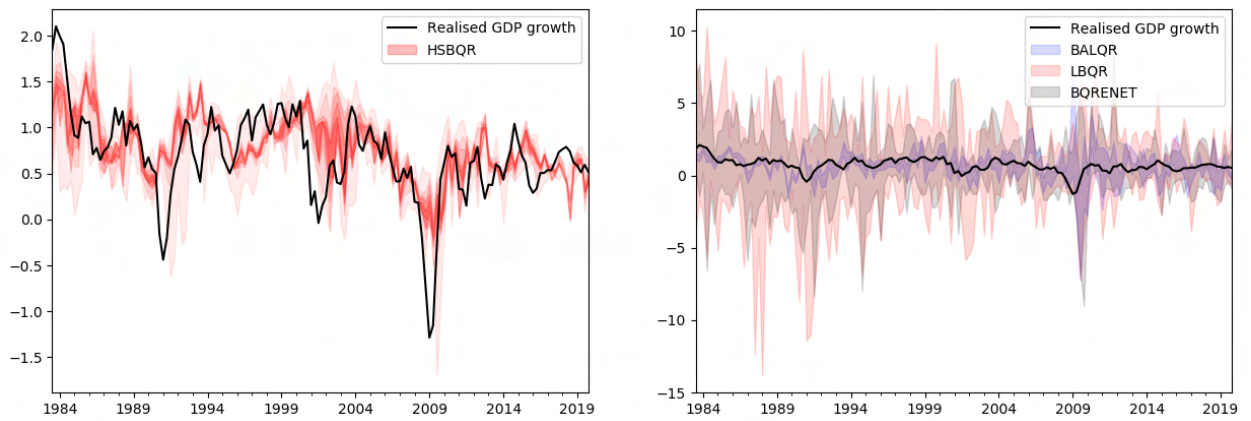


Figure C.4: Three-step-ahead forecast distributions for the L1QR, BQR, BALQR and HS-BQR. Shaded areas correspond to plots of all 19 quantiles.

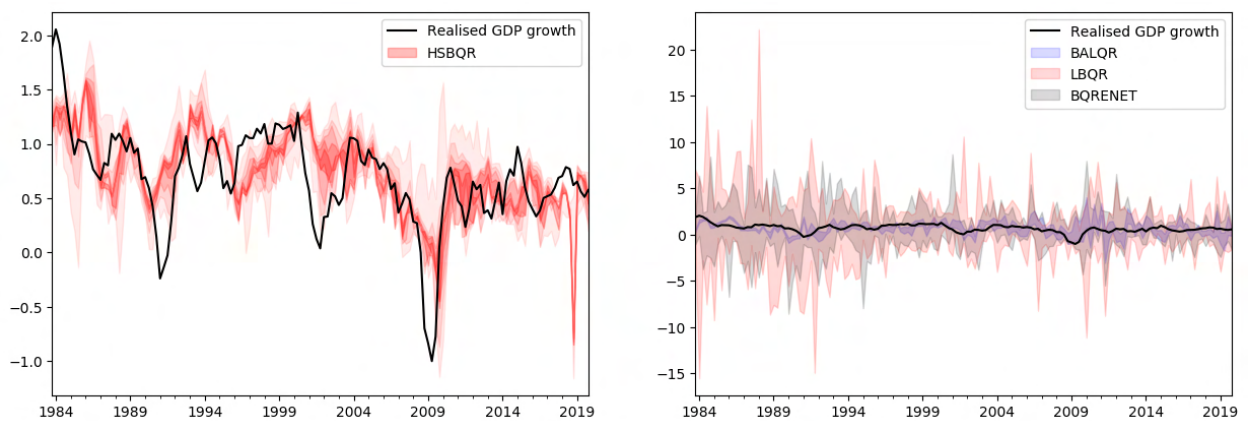


Figure C.5: Four-step-ahead forecast distributions for the L1QR, BQR, BALQR and HS-BQR. Shaded areas correspond to plots of all 19 quantiles.

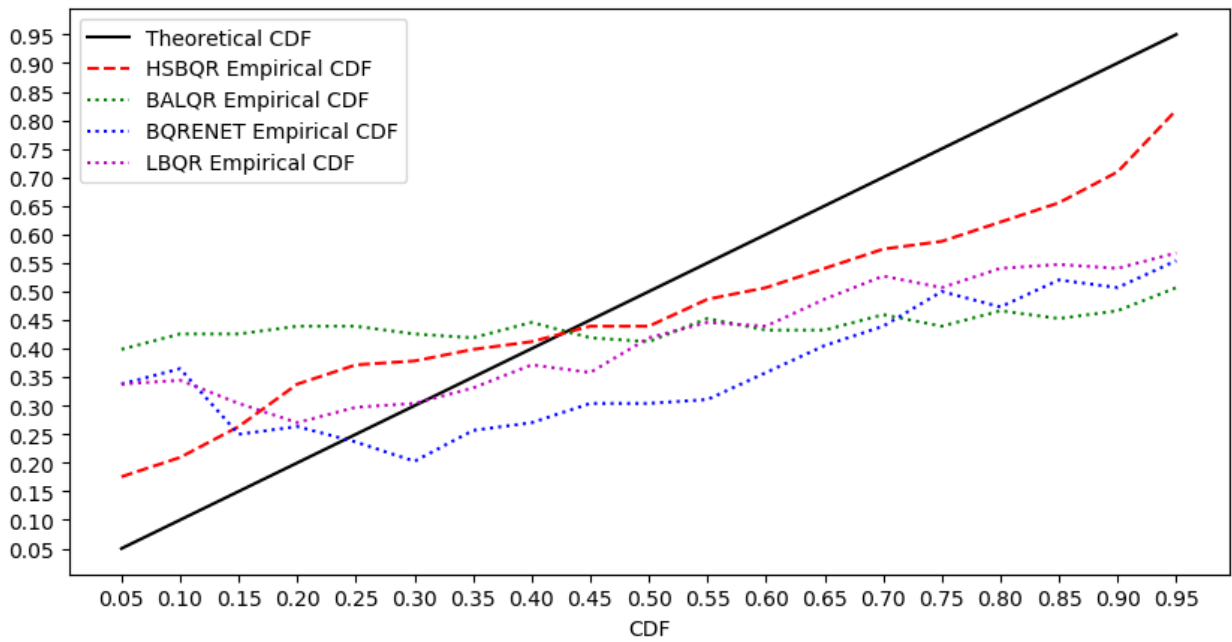


Figure C.6: One-step-ahead QQ plots of PITs for the L1QR, BQR, BALQR and HS-BQR. Theoretically optimally calibrated density has a linearly increasing QQ plot (black line).

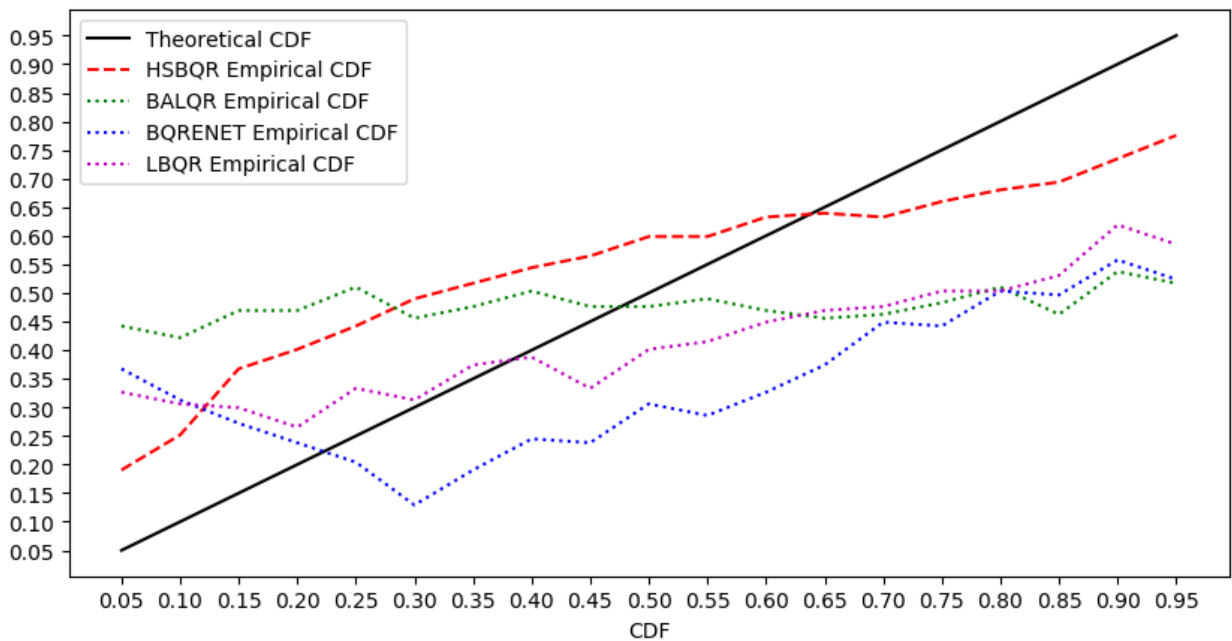


Figure C.7: Two-step-ahead QQ plots of PITs for the L1QR, BQR, BALQR and HS-BQR. Theoretically optimally calibrated density has a linearly increasing QQ plot (black line).

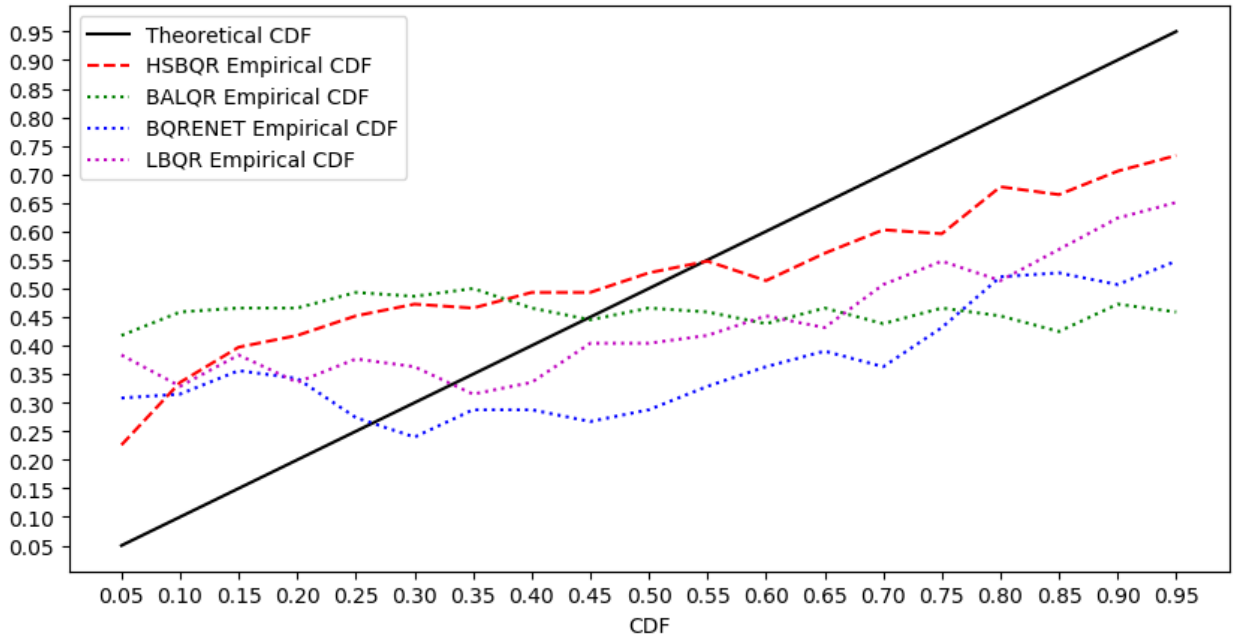


Figure C.8: Three-step-ahead QQ plots of PITs for the L1QR, BQR, BALQR and HS-BQR. Theoretically optimally calibrated density has a linearly increasing QQ plot (black line).

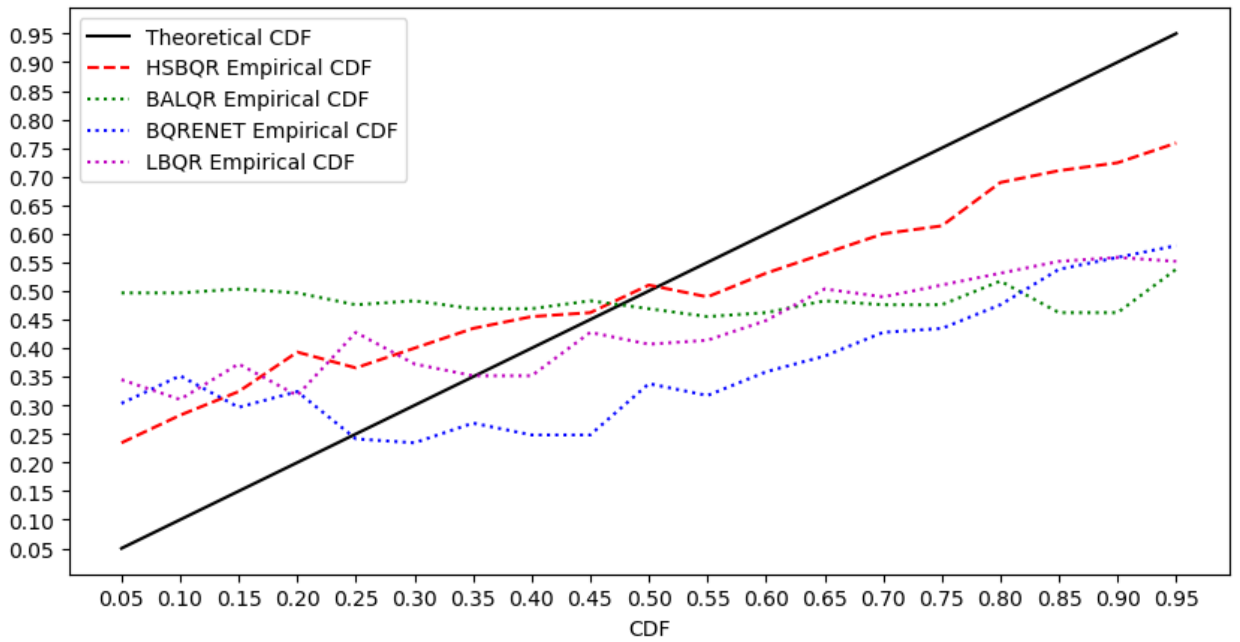


Figure C.9: Fourth-step-ahead QQ plots of PITs for the L1QR, BQR, BALQR and HS-BQR. Theoretically optimally calibrated density has a linearly increasing QQ plot (black line).

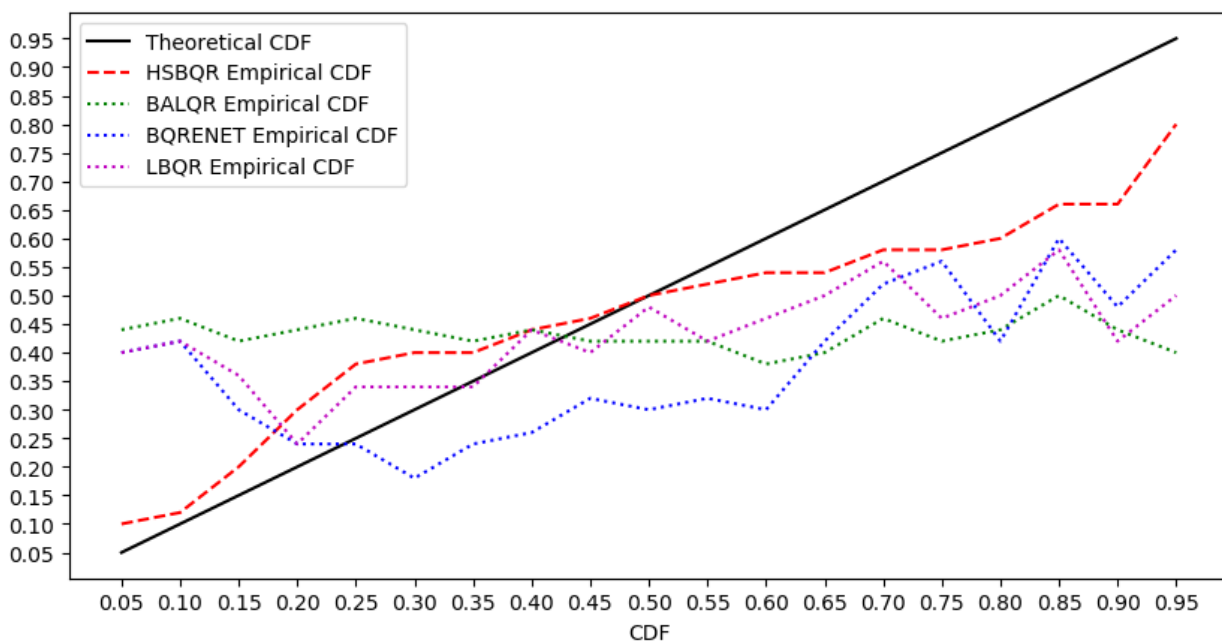


Figure C.10: One-step-ahead QQ plots of PITs for the L1QR, BQR, BALQR and HS-BQR based only on the first 50 observations. Theoretically optimally calibrated density has a linearly increasing QQ plot (black line).

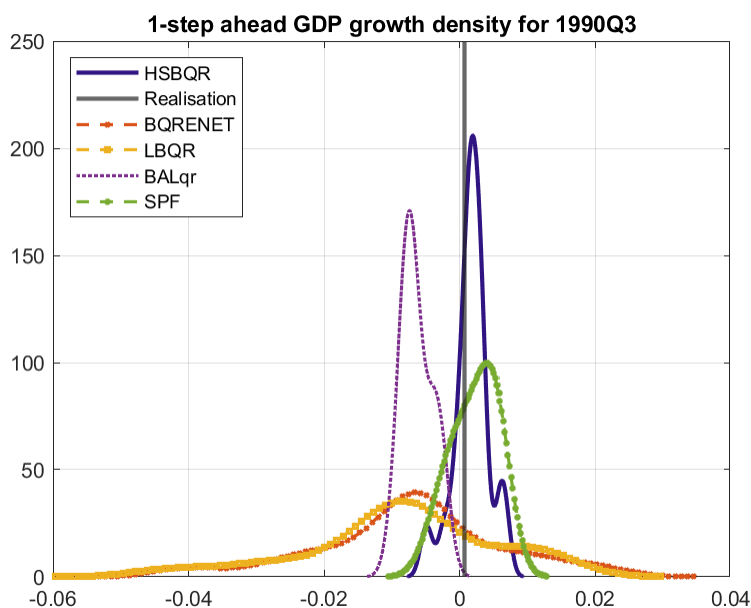


Figure C.11: Smoothed forecast densities of all competing estimators and the SPF. Densities are estimated via a Gaussian kernel of 19 equidistant forecasted quantiles. The growth realisation is marked by a vertical grey line.

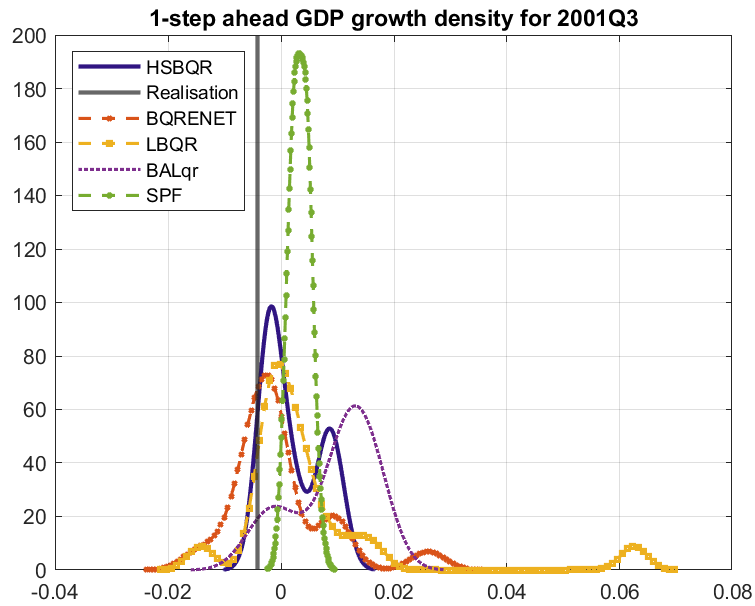


Figure C.12: Smoothed forecast densities of all competing estimators and the SPF. Densities are estimated via a Gaussian kernel of 19 equidistant forecasted quantiles. The growth realisation is marked by a vertical grey line.

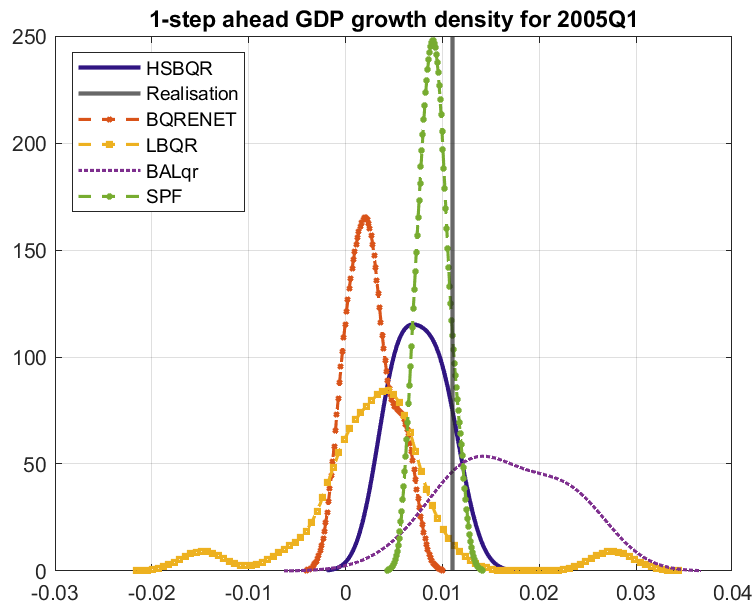


Figure C.13: Smoothed forecast densities of all competing estimators and the SPF. Densities are estimated via a Gaussian kernel of 19 equidistant forecasted quantiles. The growth realisation is marked by a vertical grey line.

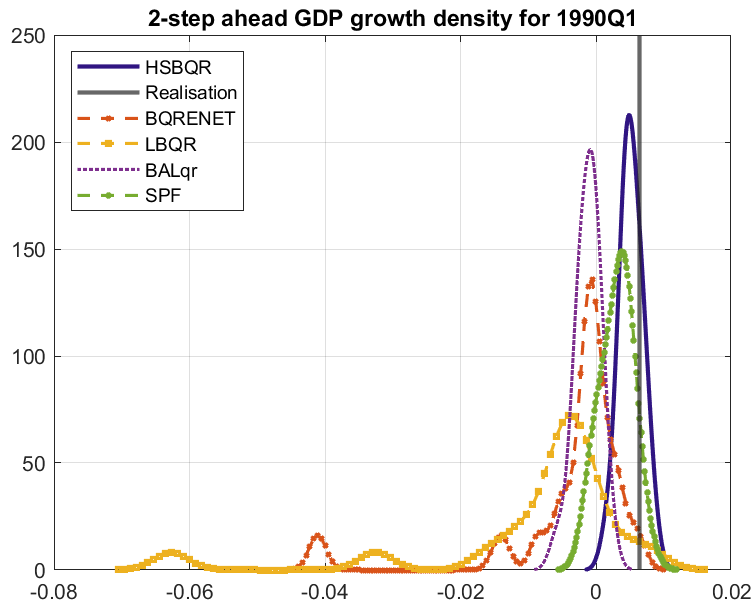


Figure C.14: Smoothed forecast densities of all competing estimators and the SPF. Densities are estimated via a Gaussian kernel of 19 equidistant forecasted quantiles. The growth realisation is marked by a vertical grey line.

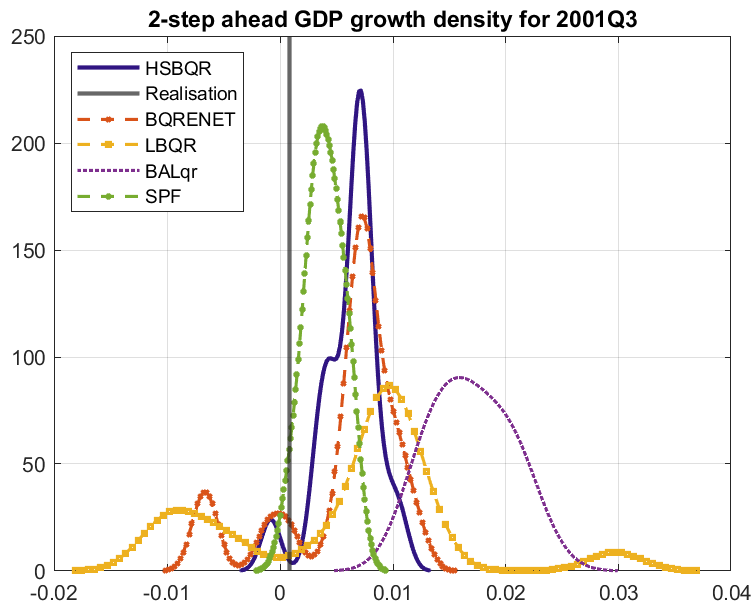


Figure C.15: Smoothed forecast densities of all competing estimators and the SPF. Densities are estimated via a Gaussian kernel of 19 equidistant forecasted quantiles. The growth realisation is marked by a vertical grey line.

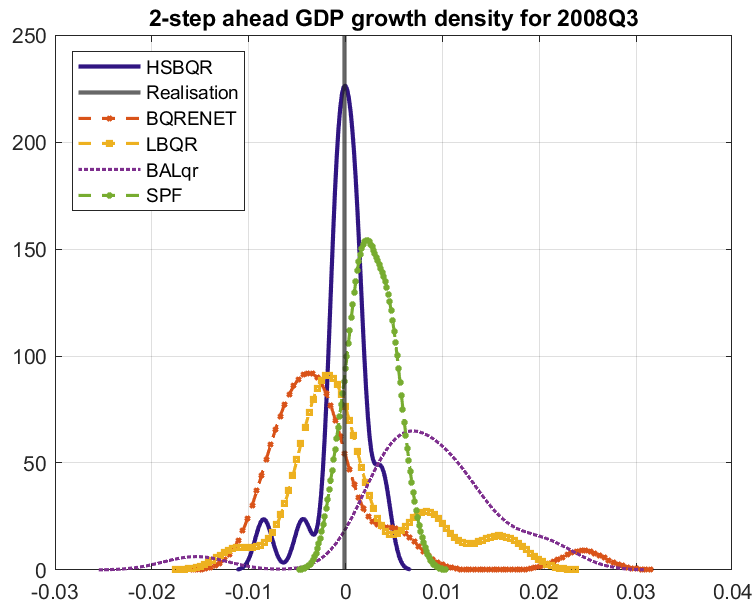


Figure C.16: Smoothed forecast densities of all competing estimators and the SPF. Densities are estimated via a Gaussian kernel of 19 equidistant forecasted quantiles. The growth realisation is marked by a vertical grey line.

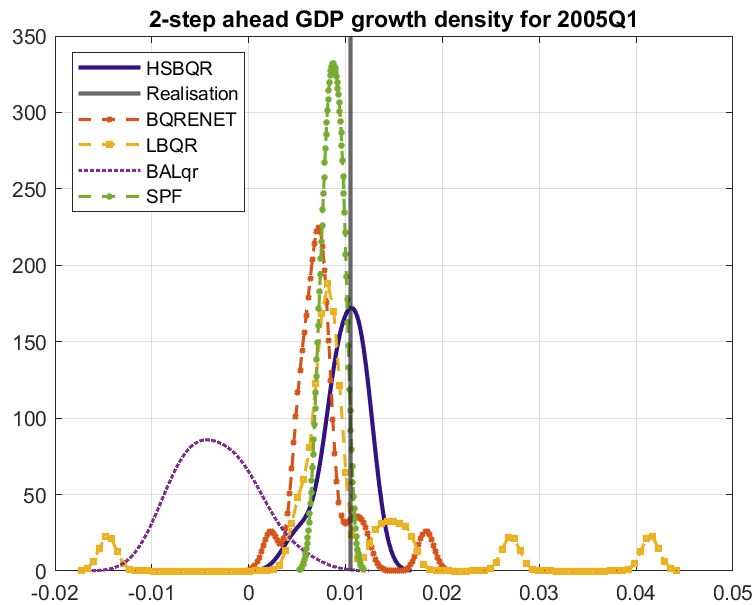


Figure C.17: Smoothed forecast densities of all competing estimators and the SPF. Densities are estimated via a Gaussian kernel of 19 equidistant forecasted quantiles. The growth realisation is marked by a vertical grey line.

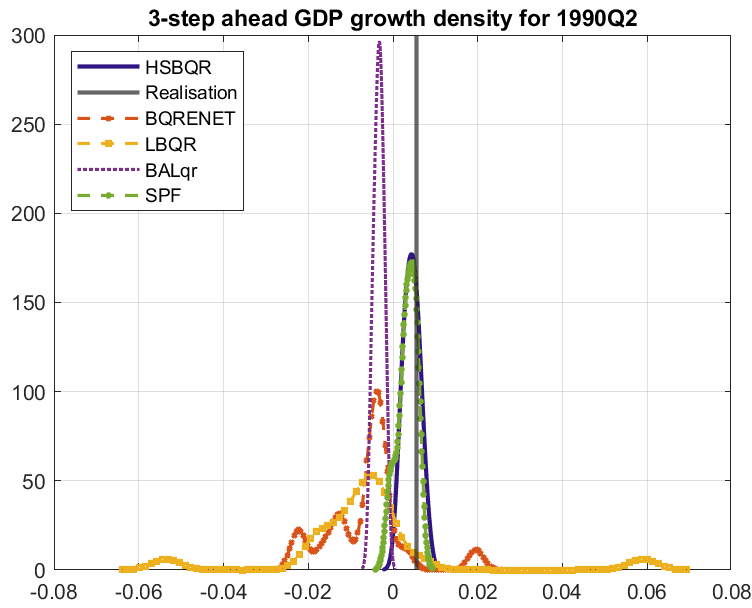


Figure C.18: Smoothed forecast densities of all competing estimators and the SPF. Densities are estimated via a Gaussian kernel of 19 equidistant forecasted quantiles. The growth realisation is marked by a vertical grey line.

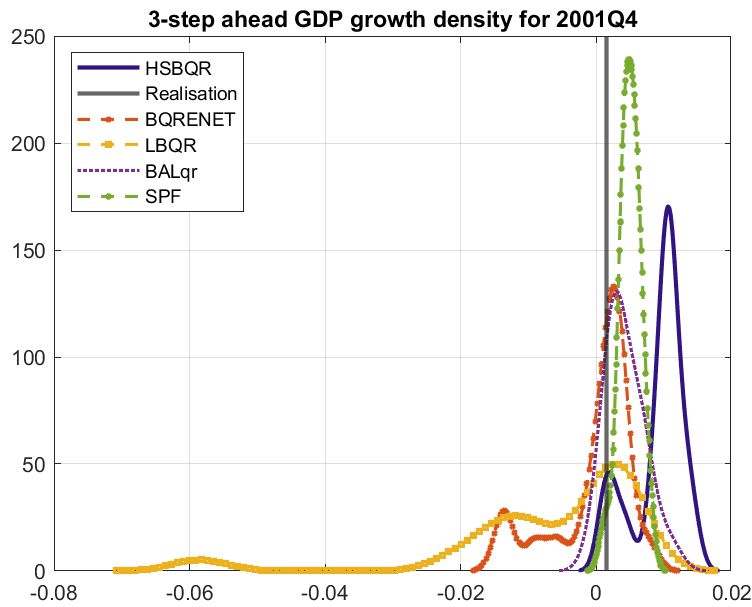


Figure C.19: Smoothed forecast densities of all competing estimators and the SPF. Densities are estimated via a Gaussian kernel of 19 equidistant forecasted quantiles. The growth realisation is marked by a vertical grey line.

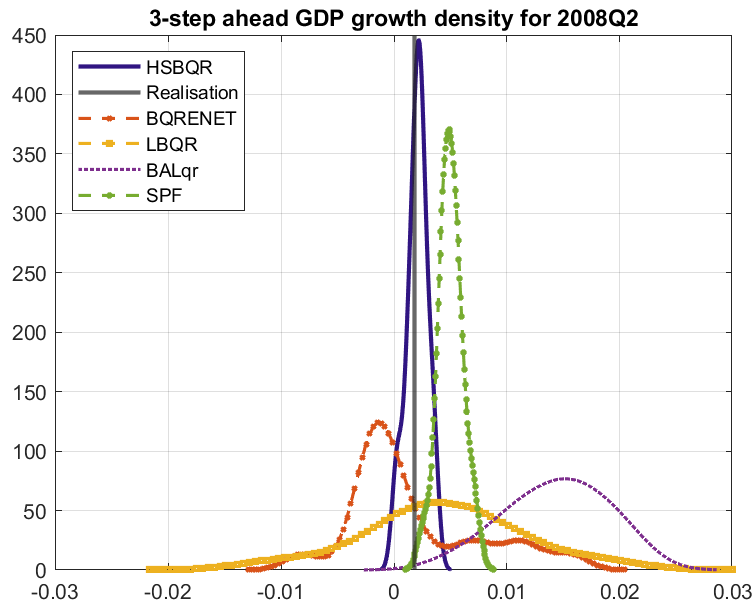


Figure C.20: Smoothed forecast densities of all competing estimators and the SPF. Densities are estimated via a Gaussian kernel of 19 equidistant forecasted quantiles. The growth realisation is marked by a vertical grey line.

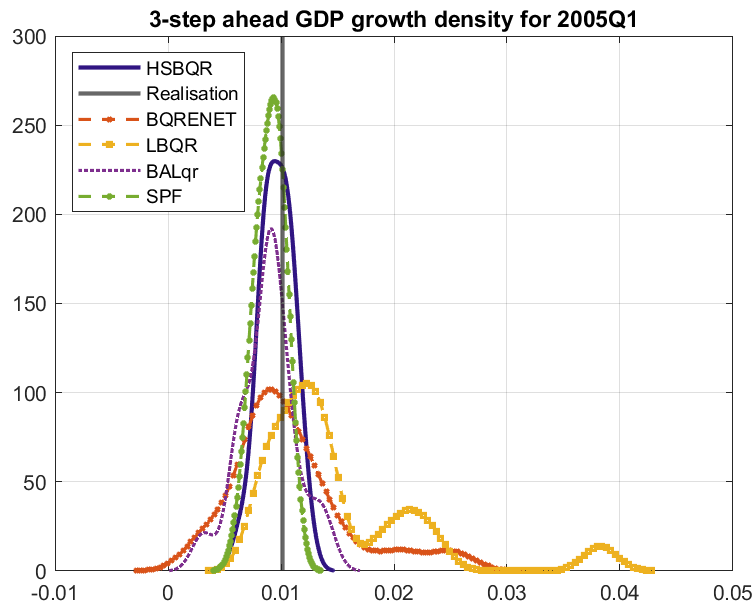


Figure C.21: Smoothed forecast densities of all competing estimators and the SPF. Densities are estimated via a Gaussian kernel of 19 equidistant forecasted quantiles. The growth realisation is marked by a vertical grey line.

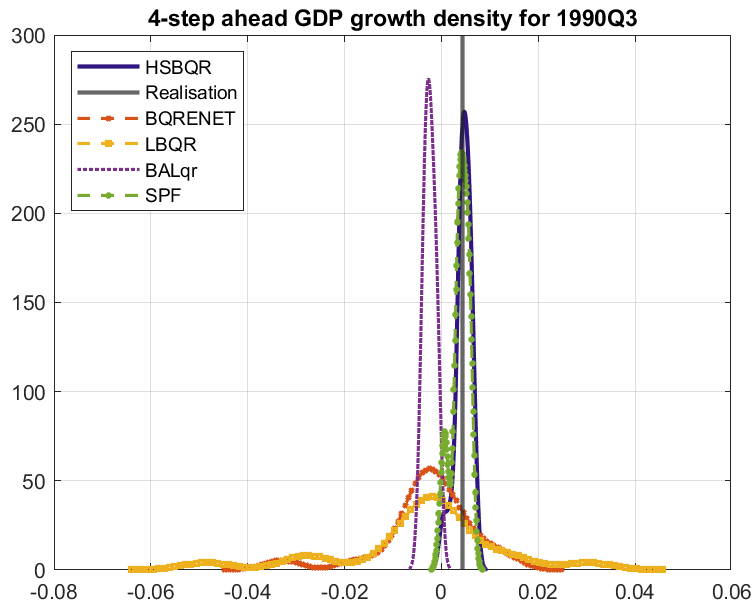


Figure C.22: Smoothed forecast densities of all competing estimators and the SPF. Densities are estimated via a Gaussian kernel of 19 equidistant forecasted quantiles. The growth realisation is marked by a vertical grey line.

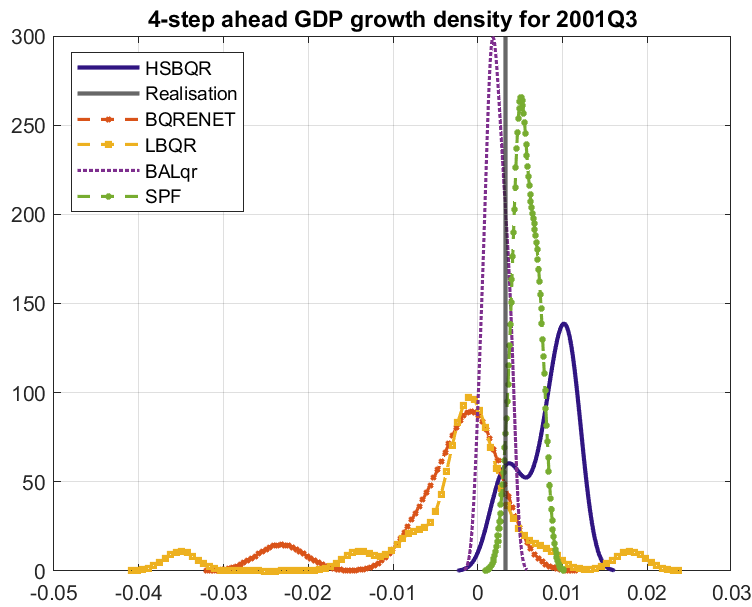


Figure C.23: Smoothed forecast densities of all competing estimators and the SPF. Densities are estimated via a Gaussian kernel of 19 equidistant forecasted quantiles. The growth realisation is marked by a vertical grey line.

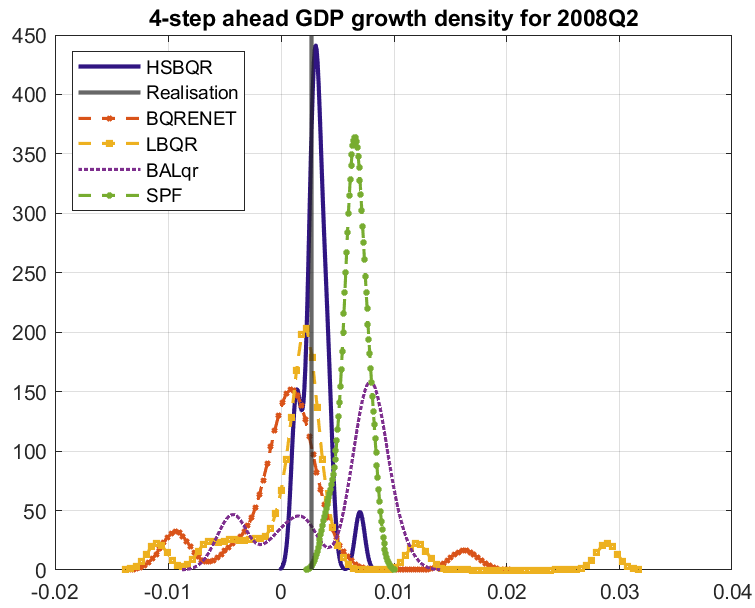


Figure C.24: Smoothed forecast densities of all competing estimators and the SPF. Densities are estimated via a Gaussian kernel of 19 equidistant forecasted quantiles. The growth realisation is marked by a vertical grey line.

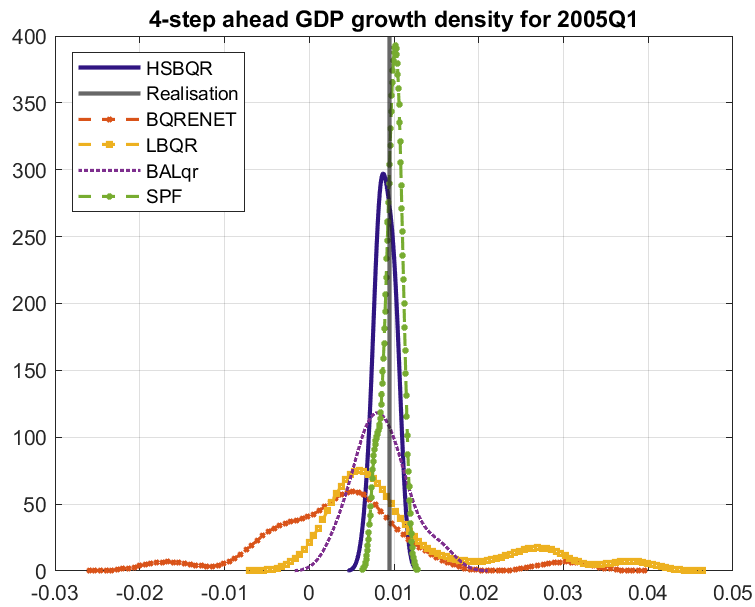


Figure C.25: Smoothed forecast densities of all competing estimators and the SPF. Densities are estimated via a Gaussian kernel of 19 equidistant forecasted quantiles. The growth realisation is marked by a vertical grey line.

Appendix D

Appendix of Chapter 4

D.1 Derivations and Further Implimentation Details

D.1.1 Expected Loss function for the BQR

In order to decouple shrinkage from sparsity, loss function (5.18) necessitates integration over the uncertainty in the latent predictions \tilde{Y} . This involves two integration steps: 1) over \tilde{Y} conditional on Υ , and 2) over Υ , conditional on \tilde{Y} . The conditional predictive likelihood, based on in-sample values of X , is:

$$y_t|\beta, Z, \sigma \sim N(x'_t\beta + \xi z_t, \tau^2 z_t\sigma) \quad (\text{D.1})$$

Note that for the sake of readability, the dependence on a given quantile is suppressed.

Integration over $\tilde{Y}|\Upsilon$

$$\begin{aligned} \mathcal{L}(\tilde{Y}, \psi) &= E_{\tilde{Y}|\bullet}[\tilde{Y}, \psi|\beta, \sigma, Z] \\ &= \zeta\|\psi\|_0 + \int T^{-1}\|X\psi - \tilde{Y}\|_2^2 p(\tilde{Y}|\beta, \sigma, Z)d\tilde{Y} \\ &= \zeta\|\psi\|_0 + T^{-1}\psi'X'X\psi - T^{-1}2\psi'X' \int \tilde{Y} p(\tilde{Y}|\beta, \sigma, Z)d\tilde{Y} \\ &\quad + T^{-1} \int \tilde{Y}'\tilde{Y} p(\tilde{Y}|\beta, \sigma, Z)d\tilde{Y} \\ &= \zeta\|\psi\|_0 + T^{-1}[\psi'X'X\psi - 2\psi X'(X\beta + \xi Z) + \Sigma_{\tilde{Y}} + \|X\beta + \xi Z\|_2^2] \\ &= \zeta\|\psi\|_0 + T^{-1}[\|X\psi - X\beta\|_2^2 - 2\psi X'(X\beta + \xi Z) + \Sigma_{\tilde{Y}} + A] \end{aligned} \quad (\text{D.2})$$

Where A is defined as $\xi^2 Z'Z + 2\beta'X'\xi Z$, and the integral of $\tilde{Y}'\tilde{Y}$ follows from:

$$\int \tilde{Y}'\tilde{Y} p(\tilde{Y}|\beta, \sigma, Z) = \sum_{t=1}^T \sigma_{\tilde{y}_t}^2 + \sum_{t=1}^T (x'_t\beta + \xi z_t)^2, \quad (\text{D.3})$$

where $\sigma_{\tilde{y}_t}^2 = \int (\tilde{y}_t - \tilde{\mu})^2 p(\tilde{y}_t | \beta, \sigma, Z)$.

Integration over β, σ, Z

$$\begin{aligned} \mathcal{L}(\psi) &= E_{\beta, \sigma, Z}[\mathcal{L}(\beta, \sigma, Z, \psi)] \\ &= \zeta \|\psi\|_0 + T^{-1} \int \|X\psi - X\beta\|_2^2 p(\beta, \sigma, Z | \tilde{Y}) d(\beta, \sigma, Z) \\ &\quad + T^{-1} \int \Sigma_{\tilde{y}} p(\beta, \sigma, Z | \tilde{Y}) d(\beta, \sigma, Z) \\ &\quad + T^{-1} A p(\beta, \sigma, Z | \tilde{Y}) d(\beta, \sigma, Z) + T^{-1} \int \psi' X' \xi Z p(\beta, \sigma, Z | \tilde{Y}) d(\beta, \sigma, Z) \end{aligned} \quad (\text{D.4})$$

Since all the terms involving $\Sigma_{\tilde{y}}$ and A don't involve ψ , they do not affect the minimisation of the loss function and can therefore be dropped.

$$\begin{aligned} &= \zeta \|\psi\|_0 + T^{-1} [\|X\psi - X\beta\|_2^2 - 2\psi' X' X \bar{\beta} + \bar{\beta}' X' X \bar{\beta} + \text{tr}(X' X \Sigma_{\bar{\beta}})] \\ &\quad - T^{-1} \psi X' \xi \int Z p(\beta, \sigma, Z | \tilde{Y}) d(\beta, \sigma, Z) \\ &= \zeta \|\psi\|_0 + T^{-1} \|X\psi - X\beta\|_2^2 + T^{-1} \text{tr}(X' X \Sigma_{\bar{\beta}}) - T^{-1} \psi' X' \xi \bar{Z}. \end{aligned} \quad (\text{D.5})$$

where $\bar{Z}_p = \left(\frac{|y_1 - x'_1 \beta_p|}{\sqrt{\xi^2 + 2\tau^2}} + \frac{\sigma\tau^2}{\xi^2 + 2\tau^2}, \dots, \frac{|y_T - x'_T \beta_p|}{\sqrt{\xi^2 + 2\tau^2}} + \frac{\sigma\tau^2}{\xi^2 + 2\tau^2} \right)'$. Entries of \bar{Z}_p follow from the definition of the expectation of the inverse Gaussian with location μ_t and scale ϑ_t : since $1/z_t \sim i\mathcal{G}(\mu_t, \vartheta_t)$, then $z_t \sim i\mathcal{G}(\frac{1}{\mu_t + \vartheta_t}, \frac{1}{\vartheta_t \mu_t} + \frac{1}{\vartheta_t^2})$ (see for example Khare and Hobert (2012)).

D.1.2 Derivation of SAVS-BQR

As described in section 5.4, instead of directly minimising (5.21), we use ℓ_1 -norm as well as adaptive penalisation akin to Zou (2006):

$$\mathcal{L}(\psi) = \underset{\psi}{\text{argmin}} \left\{ \frac{1}{2} \|X\bar{\beta} - X\psi\|_2^2 + \sum_{j=1}^K \zeta_j |\psi_j| - \psi' X' \xi \bar{Z} \right\}. \quad (\text{D.6})$$

Notice that for the first term, we use the simple fact that $\|Xa - Xb\| = \|Xb - Xa\|$, and have introduced another factor $\frac{1}{2}$ as well as dropped T^{-1} for notational convenience. We suppressed dependence of the latent quantities on quantile p for the same reason.

While the LARS algorithm (see Friedman et al. (2001)) can be applied to (D.6), we follow Ray and Bhattacharya (2018) in making use of the efficient coordinate descent algorithm introduced by Friedman et al. (2007). The coordinate descent algorithm reduces computational complexity by updating the entry in solution vector ψ_j , conditional on all other ψ , which is iterated until convergence.

For a given state of the algorithm, $\tilde{\psi}$, the objective function is recast as a function

of the j^{th} variable:

$$\tilde{\mathcal{L}}(\psi_j) = \left\{ \frac{1}{2} \|X\beta - X_{-j}\tilde{\psi}_{-j} - X_j\psi_j\|_2^2 + \sum_{k \neq j} \zeta_k |\psi| - \tilde{\psi}'_{-j} X'_{-j} \xi \bar{Z} - \psi'_j X'_j \xi \bar{Z} \right\}, \quad (\text{D.7})$$

where, likewise, X_{-j} denotes columns in X which are not the j^{th} column. Since the optimisation problem is thus broken down to a single covariate basis, the objective function (D.7) is the classical thresholding problem. Taking first order conditions for any j :

$$\begin{aligned} \frac{\partial \tilde{\mathcal{L}}_j(\psi_j)}{\partial \psi_j} &= X'_j X_j + X'_j (X_{-j} \tilde{\psi}_{-j} - X \bar{\beta}) + \zeta_j s_j + d_j = 0 \\ \psi_j &= \frac{1}{X'_j X_j} \text{sign}(X'_j \tilde{R}_j - d_j - \zeta_j s_j) \\ &= \frac{1}{X'_j X_j} \text{sign}(X'_j \tilde{R}_j - d_j) (|X'_j \tilde{R}_j - d_j| - \zeta_j)_+, \end{aligned} \quad (\text{D.8})$$

where \tilde{R}_j is the partial residual vector between $X\bar{\beta}$ and $X_{-j}\tilde{\psi}_{-j}$ at a given state and s_j is defined as the subgradient, $s_j \in \partial|\psi_j|$. Notice that we collect $-X_{-j}\xi\bar{Z}$ into d_j .

Now, since the coordinate descent algorithm is stopped after the first iteration, as suggested in Ray and Bhattacharya (2018), $R_j = \bar{\beta}_j X_j$, and therefore $X_j R_j = \bar{\beta}_j \|X\|_2^2$. Hence, by neglecting d_j as suggested in section (5.3.2), the solution to theorem 1 is achieved.

D.1.3 Sampling Algorithms

With conditional posteriors presented in section 5.2 at hand, we utilise standard Gibbs samplers to obtain draws from the posterior distributions. Collecting posterior hyperparameters for β , which will differ for each prior under investigation, in Λ_* , the dynamics of the of the Markov chain $\{\beta^m, \sigma^m, \Lambda_*^m, Z^m\}_{m=0}^\infty$ are implicitly defined through the following steps

1. Draw $z_t \sim \pi(\cdot | \beta, \sigma, \Lambda_*, \theta, \tau, Y)$ from $1/i\mathcal{G}(\bar{c}_t, \bar{d}_t)$ for all t and stack to a $T \times 1$ vector Z_{n+1}
2. Draw $\sigma_{n+1} \sim \pi(\cdot | \beta, \Lambda_*, \theta, \tau, Y, Z_{n+1})$ from $\mathcal{G}^{-1}(\bar{a}, \bar{b})$
3. Draw $\beta_{n+1} \sim \pi(\cdot | \sigma_{n+1}, \Lambda_*, \theta, \tau, Y, Z_{n+1})$ from $N(\bar{\beta}, \bar{\Lambda}_*)$
4. Draw Λ_{n+1} according to each prior in section 5.2
5. Iterate (1-4) until convergence is achieved

Note that for the horseshoe prior, we use a slice sampler to sample the elements of Λ_* , as suggested in Kohns and Szendrei (2020).

Fast BQR Sampler

The most costly operation in the Gibbs sampler in (D.1.3), is the inversion of the possibly large dimensional posterior covariance of the regression coefficients, step 4. To speed up computation, we make use of the fast BQR sampler proposed in Kohns and Szendrei (2020):

As derived section (5.7), using the scale mixture representation in, the conditional posterior of β given all other parameters can be written as:

$$\beta|\bullet \sim N(A^{-1}X'\Sigma y, A^{-1}), \quad A = (X'\Sigma X + \Lambda_*^{-1}) \quad (\text{D.9})$$

Suppose, we want to sample from $N_K(\mu, \Theta)$, where

$$\Theta = (\Phi'\Phi + D^{-1})^{-1}, \quad \mu = \Theta\Phi'(\alpha - \xi Z). \quad (\text{D.10})$$

Assume $D \in \mathbb{R}^{K \times K}$ is a positive definitive matrix and diagonal for simplicity, $\phi \in \mathbb{R}^{T \times K}$, and $\alpha \in \mathbb{R}^{T \times 1}$. Then (5.7) is recovered when setting $\Phi = U^{1/2}X$, $D = \Lambda_*$ and $\alpha = U^{1/2}y$. An exact algorithm to sample from (5.7) is thus given by:

Algorithm 2. *Fast HS-BQR sampler*

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

Suppose ς is obtained through algorithm 1. Then $\varsigma \sim N(\mu, \Theta)$. For a proof, we refer to Kohns and Szendrei (2020).

Conditional Mean Models

Assuming a Gaussian likelihood for linear model $y = X\beta + \epsilon$, $\epsilon \sim N(0, \sigma^2 I_T)$, the prior for regression coefficients, β , is assumed to be multivariate normal, and the error variance, σ^2 , inverse-Gamma:

$$\begin{aligned} p(\beta|V) &\sim N(0, V) \\ p(V) &\sim f \\ p(\sigma^2) &\sim \mathcal{G}^{-1}(\underline{a}, \underline{b}). \end{aligned} \quad (\text{D.11})$$

The horseshoe, SSVS and lasso priors differ in how the hierarchical prior variance on β , $p(V)$, is specified. For all priors, we set $\underline{a} = \underline{b} = 0.1$.

The conditional posteriors due to conditional conjugacy are:

$$\begin{aligned}\beta|\bullet &\sim N(\bar{\beta}, \bar{V}) \\ \sigma^2|\bullet &\sim \mathcal{G}^{-1}(\bar{a}/2, \bar{b}/2) \\ \bar{V}|\bullet &\sim g,\end{aligned}\tag{D.12}$$

where $\bar{a} = \underline{a} + T$ and $\bar{b} = \underline{b} + \frac{(y-X\beta)'(y-X\beta)}{T-K} + \bar{\beta}'[V + (X'X)^{-1}]^{-1}\bar{\beta}$. The posterior of the variance hyperparameters and of the regression parameters are prior specific and detailed below. These conditional posteriors are sampled in a standard Gibbs sampler with 5000 burnin draws and 5000 retained MCMC samples.

Horseshoe

For the horseshoe prior, set $V = \sigma^2\vartheta^2 \text{diag}(\varphi_1^2, \dots, \varphi_K^2)$, where

$$\begin{aligned}\vartheta &\sim C_+(0, 1) \\ \varphi_j &\sim C_+(0, 1)\forall j.\end{aligned}\tag{D.13}$$

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

$$\begin{aligned}\bar{\beta} &= \bar{V}^{-1}X'y, \\ \bar{V} &= \sigma^2(X'X + V^{-1})\end{aligned}\tag{D.14}$$

Instead of computing the large dimensional inverse A^{-1} , we rely on a data augmentation technique introduced by Bhattacharya et al. (2016).

SSVS

For the SSVS prior, we use the same hierarchy as presented in (5.16), which is printed below for convenience:

$$\begin{aligned}\beta_{j,p}|\gamma_j, \varphi_j &\sim (1 - \gamma_j)N(0, c\varphi_j^2) + \gamma_j N(0, \varphi_j^2)\forall j \in \{1, \dots, K\} \\ \varphi_j^2 &\sim \mathcal{G}^{-1}(a_2, b_2) \\ \gamma_j|\pi_0 &\sim \text{Bern}(\pi_0) \\ \pi_0 &\sim \mathcal{B}(a_3, b_3),\end{aligned}\tag{D.15}$$

where $V = \text{diag}(\varphi_1^2, \dots, \varphi_K^2)$ for all j if $\gamma_j = 1$, and $\text{diag}(c\varphi_1^2, \dots, c\varphi_k^2)$ otherwise. 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 on 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):

$$\begin{aligned}
\bar{\beta} &= \bar{V}^{-1} X' y / \sigma^2 \\
\bar{V} &= (X' X / \sigma^2 + V^{-1})^{-1}, \quad V = \text{diag}(\varphi_j \gamma_j) \\
\gamma_j | \bullet &\sim N(1 - \pi_0) N(\beta_j | 0, c \times \varphi_j^2) I_{\gamma_j=0} + \pi_0 N(\beta_j | 0, \varphi_j^2) I_{\gamma_j=1} \\
\pi_0 | \bullet &\sim \mathcal{B}(a_3 + n_1, b_3 + K - n_1), \text{ where } n_1 = \sum_j I_{\gamma_j=1}
\end{aligned} \tag{D.16}$$

To be agnostic about the degree of sparsity, we set $a_3 = b_3 = 0$ as for the quantile SSVS model.

Lasso

Similar to the SSVS, the lasso prior is the same as in (5.11):

$$\begin{aligned}
\beta_j | \varphi &\sim N(0, \varphi_j), \\
\varphi_j | \phi &\sim \exp\left(\frac{\phi}{2}\right) \\
\phi &\sim \mathcal{G}(a_1, b_1)
\end{aligned} \tag{D.17}$$

where $\Lambda_* = \text{diag}(\vartheta_1, \dots, \vartheta_K)$. The conditional posteriors for the hyperparameters are standard:

$$\begin{aligned}
p(\vartheta_j^{-1} | \bullet) &\sim i\mathcal{G}\left(\sqrt{\frac{\phi}{\beta_{j,p}^2}}, \phi\right) \\
p(\phi | \bullet) &\sim \mathcal{G}\left(K + a_1, \frac{1}{2} \sum_{j=1}^K \vartheta_j + b_1\right) \\
\bar{\beta} &= \bar{V}^{-1} X' y / \sigma^2 \\
\bar{V} &= (X' X / \sigma^2 + V^{-1})^{-1}, \quad V = \text{diag}(\vartheta_1, \dots, \vartheta_K)
\end{aligned} \tag{D.18}$$

D.2 Supplementary Material on the Empirical Application

	MSFE	LPDS	CRPS	qwCRPS	MSFE	LPDS	CRPS	qwCRPS
	h=1				h=2			
<i>ABG_{BQR}</i>	0.504	-11.607	0.313	1.049	0.404	-7.870	0.254	0.891
<i>ABG_{BQR-Skt}</i>	0.519	-1.727	0.301	1.026	0.426	-4.102	0.277	0.970
<i>SSVSBQR</i>	0.539*	-0.889**	0.320***	1.061	0.435**	-0.607	0.221***	0.746**
<i>HSBQR</i>	0.542**	-3.695	0.313***	0.952*	0.473**	-5.101	0.236***	0.776**
<i>HSBQR_{BIC}</i>	0.522	-3.020	0.300***	0.918**	0.428**	-5.775*	0.239***	0.779***
<i>HSBQR_{SAVS}</i>	0.537	-2.838	0.304***	0.925*	0.416	-3.959	0.259***	0.847
<i>LBQR</i>	0.570	-1.313	0.289***	0.949	0.480	-0.760*	0.229***	0.748***
<i>LBQR_{BIC}</i>	0.544	-1.316	0.289***	0.949	0.450	-0.776*	0.230***	0.750***
<i>LBQR_{SAVS}</i>	0.544	-2.021	0.301**	0.999	0.449	-1.219*	0.245***	0.808
	h=3				h=4			
<i>ABG_{BQR}</i>	0.378	-6.825	0.237	0.847	0.351	-4.387	0.235	0.844
<i>ABG_{BQR-Skt}</i>	0.397	-1.006	0.249	0.857	0.367	-1.252	0.251	0.875
<i>SSVSBQR</i>	0.431**	-0.542	0.211***	0.723	0.392*	-0.657*	0.214***	0.710
<i>HSBQR</i>	0.437**	-5.224	0.231***	0.777	0.392**	-5.258*	0.224***	0.731
<i>HSBQR_{BIC}</i>	0.400*	-5.958**	0.230***	0.775	0.377*	-4.815**	0.220***	0.723
<i>HSBQR_{SAVS}</i>	0.395	-4.714	0.240***	0.810	0.380	-3.568	0.216***	0.729
<i>LBQR</i>	0.431*	-0.620	0.222***	0.743**	0.413*	-0.667	0.213***	0.716**
<i>LBQR_{BIC}</i>	0.422*	-0.630	0.223***	0.745**	0.387*	-0.675	0.214***	0.718**
<i>LBQR_{SAVS}</i>	0.421*	-0.964	0.233***	0.797	0.386*	-1.557	0.234***	0.804

Table D.1: Forecast Evaluation Results (Skewed-t distribution)

	0.05	0.25	0.5	0.75	0.95	0.05	0.25	0.5	0.75	0.95	0.05	0.25	0.5	0.75	0.95	0.05	0.25	0.5	0.75	0.95	
	y_1					y_2					y_3					y_4					
Coefficient Bias																					
T=100																					
Sparse																					
<i>HSBQRZ</i>	0.121	0.092	0.086	0.098	0.126	0.172	0.099	0.088	0.097	0.176	0.156	0.098	0.077	0.094	0.148	0.128	0.095	0.079	0.094	0.127	
<i>HSBQRZ_{SAVS}</i>	0.110	0.072	0.064	0.081	0.116	0.165	0.081	0.066	0.080	0.169	0.146	0.089	0.062	0.086	0.138	0.119	0.085	0.063	0.085	0.119	
<i>HSBQRZ_{BIC}</i>	0.103	0.062	0.058	0.072	0.107	0.159	0.067	0.056	0.072	0.163	0.141	0.087	0.060	0.085	0.132	0.117	0.084	0.061	0.084	0.116	
Block																					
<i>HSBQRZ</i>	0.250	0.247	0.252	0.254	0.244	0.260	0.240	0.244	0.250	0.272	0.289	0.293	0.291	0.285	0.277	0.257	0.266	0.268	0.263	0.256	
<i>HSBQRZ_{SAVS}</i>	0.252	0.242	0.246	0.248	0.243	0.262	0.236	0.239	0.245	0.272	0.290	0.287	0.283	0.279	0.276	0.260	0.262	0.262	0.258	0.256	
<i>HSBQRZ_{BIC}</i>	0.259	0.246	0.247	0.247	0.247	0.271	0.241	0.244	0.248	0.276	0.292	0.286	0.283	0.278	0.277	0.262	0.262	0.262	0.258	0.258	
T=500																					
Sparse																					
<i>HSBQRZ</i>	0.099	0.073	0.062	0.073	0.103	0.152	0.074	0.061	0.075	0.152	0.140	0.066	0.048	0.067	0.135	0.113	0.065	0.047	0.066	0.111	
<i>HSBQRZ_{SAVS}</i>	0.086	0.053	0.038	0.054	0.089	0.145	0.057	0.037	0.059	0.144	0.127	0.058	0.026	0.058	0.123	0.100	0.056	0.026	0.057	0.098	
<i>HSBQRZ_{BIC}</i>	0.076	0.045	0.029	0.047	0.077	0.140	0.050	0.030	0.054	0.137	0.106	0.061	0.025	0.062	0.106	0.084	0.058	0.024	0.059	0.085	
Block																					
<i>HSBQRZ</i>	0.143	0.101	0.094	0.103	0.144	0.174	0.104	0.091	0.103	0.176	0.187	0.098	0.082	0.096	0.184	0.156	0.101	0.085	0.099	0.154	
<i>HSBQRZ_{SAVS}</i>	0.139	0.089	0.081	0.092	0.138	0.173	0.093	0.077	0.092	0.171	0.185	0.091	0.072	0.087	0.181	0.152	0.092	0.074	0.091	0.150	
<i>HSBQRZ_{BIC}</i>	0.146	0.096	0.085	0.099	0.151	0.180	0.099	0.080	0.097	0.183	0.190	0.091	0.072	0.088	0.186	0.156	0.092	0.074	0.091	0.154	
MCC																					
T=100																					
Sparse																					
<i>HSBQRZ_{SAVS}</i>	0.392	0.535	0.555	0.548	0.509	0.383	0.516	0.552	0.556	0.504	0.338	0.446	0.446	0.430	0.324	0.382	0.453	0.433	0.424	0.365	
<i>HSBQRZ_{BIC}</i>	0.555	0.725	0.780	0.772	0.713	0.561	0.722	0.793	0.791	0.707	0.423	0.496	0.540	0.464	0.405	0.446	0.485	0.523	0.448	0.426	
Block																					
<i>HSBQRZ_{SAVS}</i>	0.386	0.412	0.420	0.411	0.393	0.405	0.429	0.439	0.421	0.398	0.336	0.342	0.344	0.351	0.345	0.384	0.392	0.386	0.394	0.382	
<i>HSBQRZ_{BIC}</i>	0.420	0.445	0.457	0.452	0.435	0.435	0.457	0.471	0.462	0.441	0.372	0.371	0.350	0.376	0.383	0.413	0.405	0.372	0.405	0.411	
T=500																					
Sparse																					
<i>HSBQRZ_{SAVS}</i>	0.397	0.540	0.567	0.543	0.462	0.389	0.568	0.576	0.574	0.471	0.362	0.679	0.641	0.653	0.408	0.398	0.668	0.648	0.654	0.434	
<i>HSBQRZ_{BIC}</i>	0.681	0.898	0.904	0.898	0.840	0.702	0.905	0.906	0.903	0.850	0.659	0.783	0.836	0.792	0.665	0.709	0.785	0.827	0.785	0.697	
Block																					
<i>HSBQRZ_{SAVS}</i>	0.691	0.804	0.810	0.807	0.691	0.707	0.809	0.820	0.815	0.710	0.613	0.837	0.865	0.850	0.617	0.676	0.821	0.855	0.844	0.677	
<i>HSBQRZ_{BIC}</i>	0.779	0.946	0.947	0.943	0.828	0.781	0.946	0.951	0.952	0.839	0.714	0.889	0.907	0.899	0.717	0.767	0.861	0.883	0.875	0.765	
Hit rate																					
T=100																					
Sparse																					
<i>HSBQRZ_{SAVS}</i>	0.578	0.817	0.842	0.835	0.768	0.567	0.790	0.828	0.822	0.762	0.538	0.608	0.612	0.565	0.511	0.540	0.618	0.608	0.571	0.518	
<i>HSBQRZ_{BIC}</i>	0.481	0.640	0.707	0.703	0.690	0.495	0.642	0.728	0.724	0.684	0.479	0.546	0.484	0.507	0.460	0.501	0.586	0.509	0.536	0.481	
Block																					
<i>HSBQRZ_{SAVS}</i>	0.554	0.652	0.685	0.671	0.583	0.570	0.665	0.691	0.672	0.583	0.549	0.631	0.650	0.639	0.556	0.566	0.649	0.667	0.655	0.567	
<i>HSBQRZ_{BIC}</i>	0.358	0.414	0.443	0.439	0.388	0.368	0.423	0.444	0.436	0.388	0.454	0.570	0.632	0.581	0.461	0.499	0.624	0.686	0.632	0.498	
T=500																					
Sparse																					
<i>HSBQRZ_{SAVS}</i>	0.769	0.956	0.966	0.951	0.924	0.745	0.950	0.957	0.963	0.916	0.818	0.935	0.949	0.920	0.830	0.834	0.940	0.956	0.927	0.845	
<i>HSBQRZ_{BIC}</i>	0.651	0.831	0.853	0.843	0.763	0.632	0.846	0.862	0.848	0.786	0.656	0.889	0.887	0.868	0.653	0.658	0.905	0.926	0.891	0.655	
Block																					
<i>HSBQRZ_{SAVS}</i>	0.963	0.999	0.998	0.998	0.969	0.966	0.998	1.000	0.998	0.972	0.932	0.986	0.995	0.989	0.929	0.954	0.988	0.995	0.990	0.953	
<i>HSBQRZ_{BIC}</i>	0.905	0.982	0.989	0.980	0.875	0.918	0.983	0.994	0.985	0.876	0.852	0.982	0.993	0.985	0.854	0.915	0.986	0.994	0.988	0.913	

Table D.2: Nuisance Parameter correction results for the HSBQR

Appendix E

Appendix of Chapter 5

E.1 Supplementary Material on the Empirical Applications

E.1.1 Implicit Model Sizes

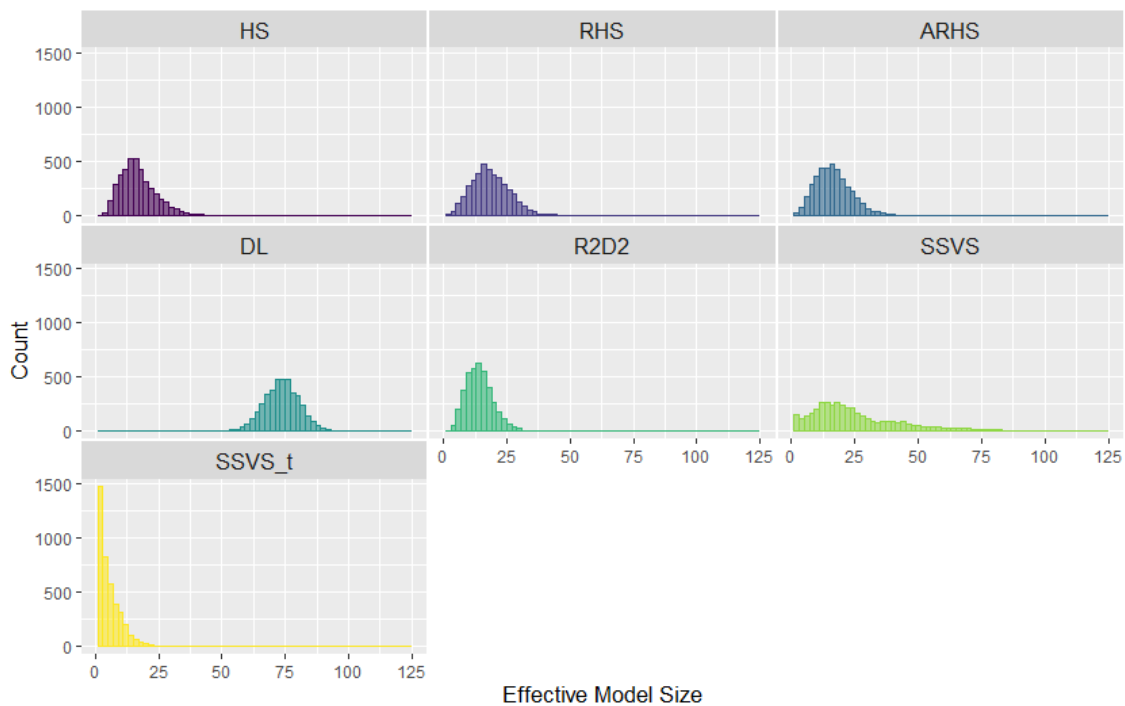


Figure E.1: Macro 1: Implicit model sizes for the HS, RHS, A-RHS, DL, R2D2 priors and the posterior model size distributions for the SSVS and SSVS-t model.

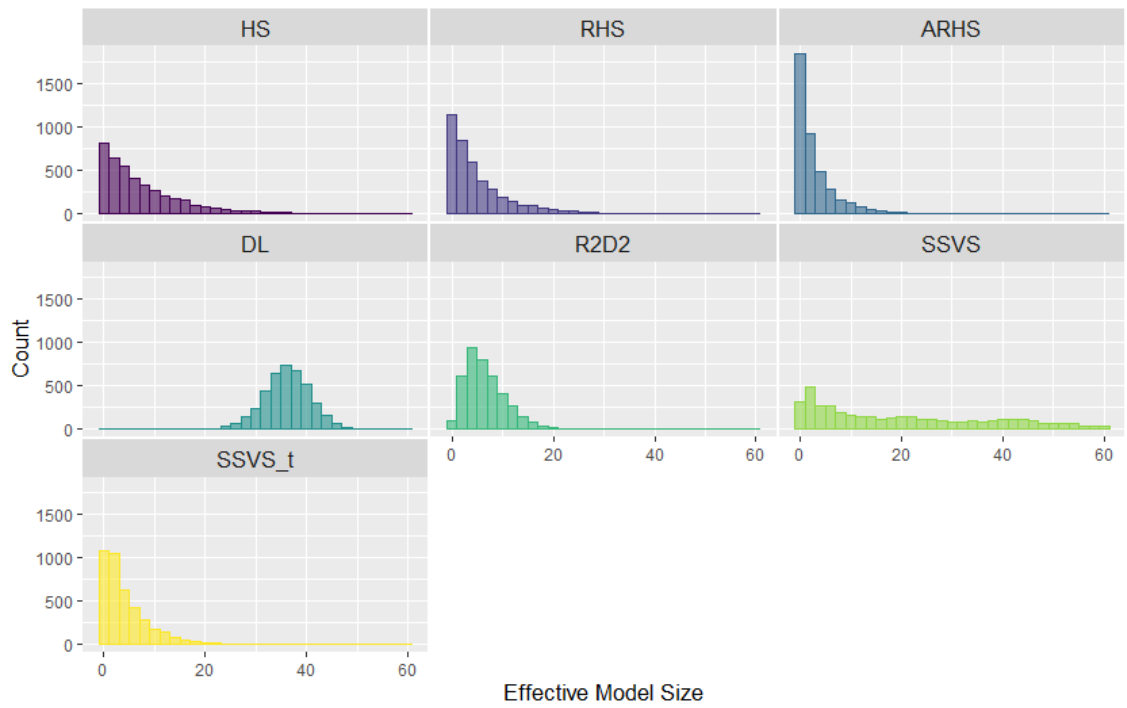


Figure E.2: Macro 2: Implicit model sizes for the HS, RHS, A-RHS, DL, R2D2 priors and the posterior model size distributions for the SSVS and SSVS-t model.



Figure E.3: Finance 1: Implicit model sizes for the HS, RHS, A-RHS, DL, R2D2 priors and the posterior model size distributions for the SSVS and SSVS-t model.

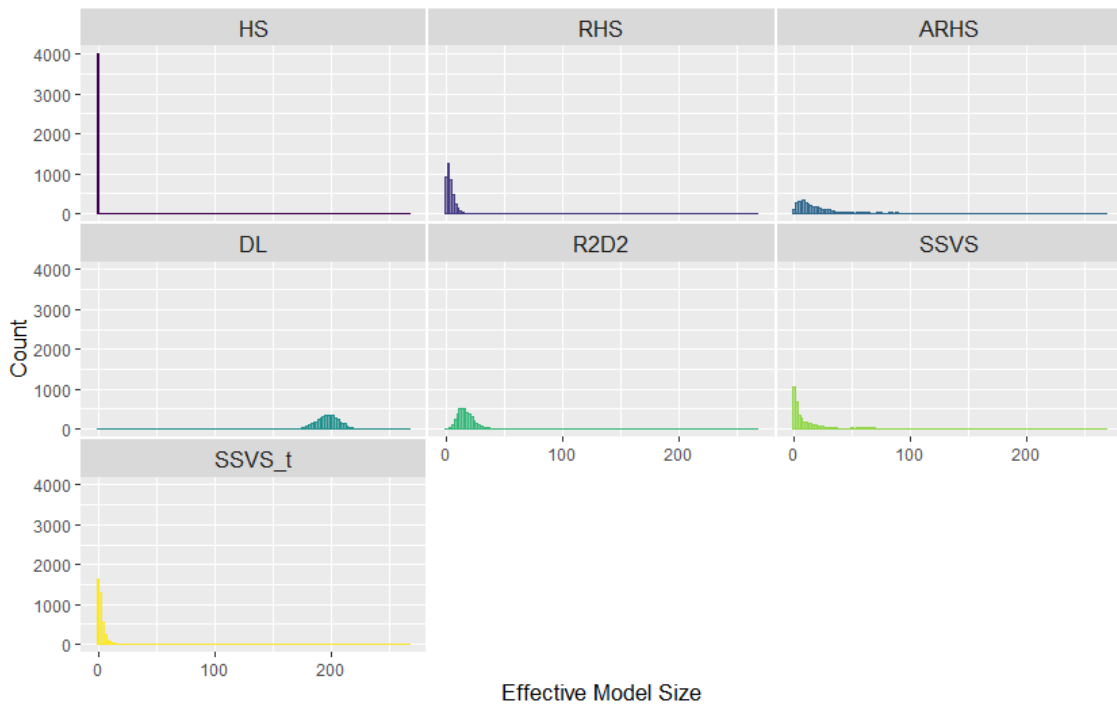


Figure E.4: Micro 1: Implicit model sizes for the HS, RHS, A-RHS, DL, R2D2 priors and the posterior model size distributions for the SSVS and SSVS-t model.



Figure E.5: Micro 2: Implicit model sizes for the HS, RHS, A-RHS, DL, R2D2 priors and the posterior model size distributions for the SSVS and SSVS-t model.

E.1.2 SAVS Model Sizes

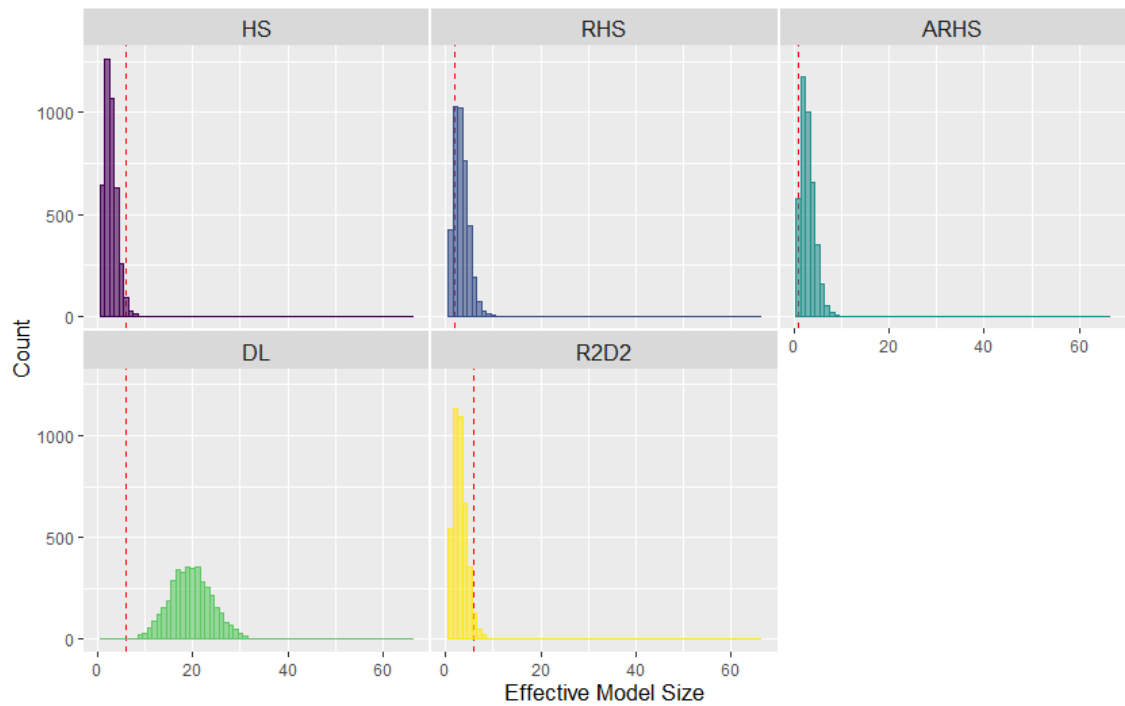


Figure E.6: Macro 1: SAVS posterior model sizes for the HS, RHS, A-RHS, DL, R2D2 priors and the posterior model size distributions for the SSVS and SSVS-t model. The red-dashed line indicated the model size chosen by Projpred

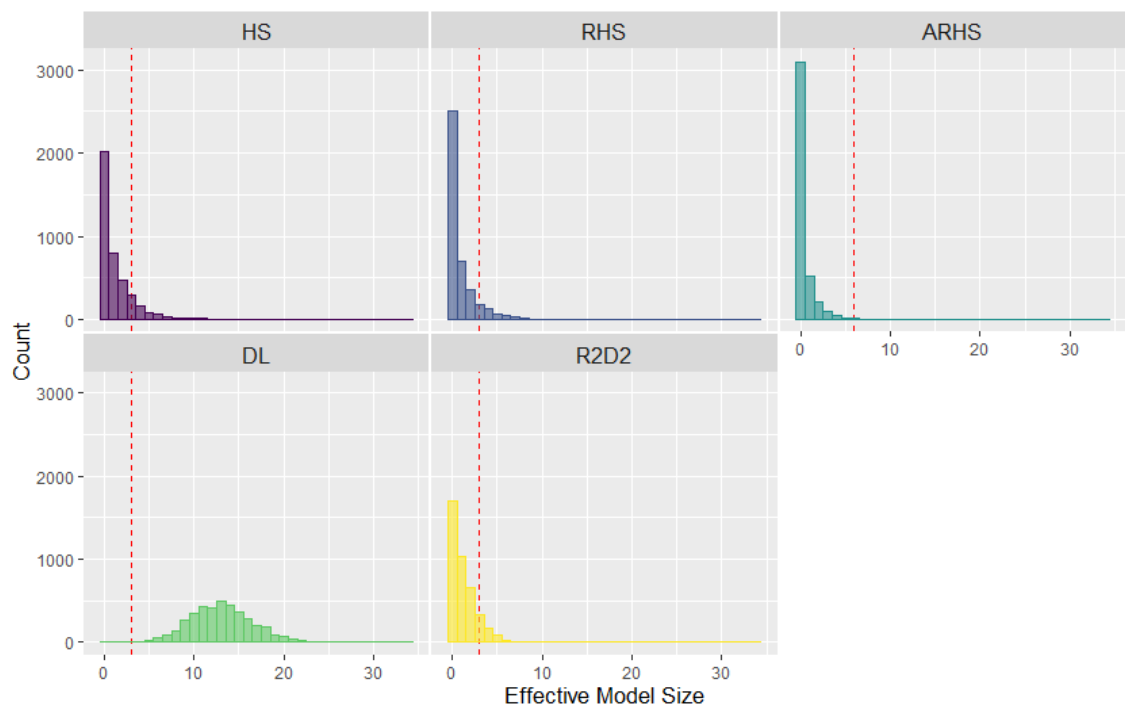


Figure E.7: Macro 2: SAVS posterior model sizes for the HS, RHS, A-RHS, DL, R2D2 priors and the posterior model size distributions for the SSVS and SSVS-t model. The red-dashed line indicated the model size chosen by Projpred

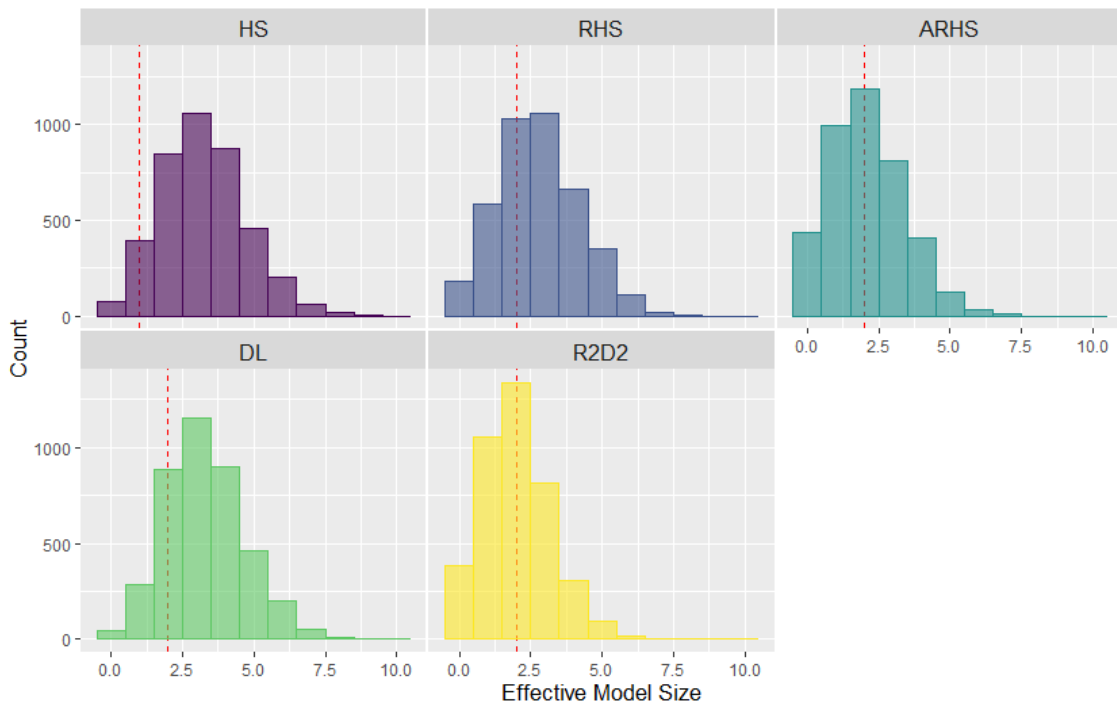


Figure E.8: Finance 1: SAVS posterior model sizes for the HS, RHS, A-RHS, DL, R2D2 priors and the posterior model size distributions for the SSVS and SSVS-t model. The red-dashed line indicated the model size chosen by Projpred

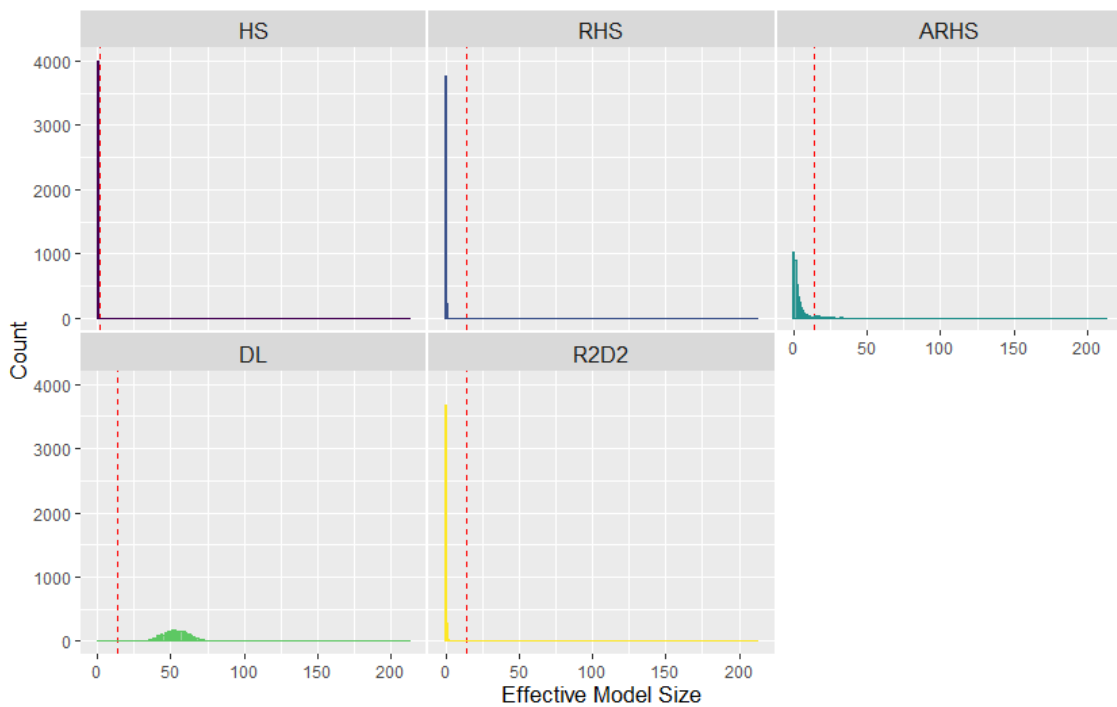


Figure E.9: Micro 1: SAVS posterior model sizes for the HS, RHS, A-RHS, DL, R2D2 priors and the posterior model size distributions for the SSVS and SSVS-t model. The red-dashed line indicated the model size chosen by Projpred

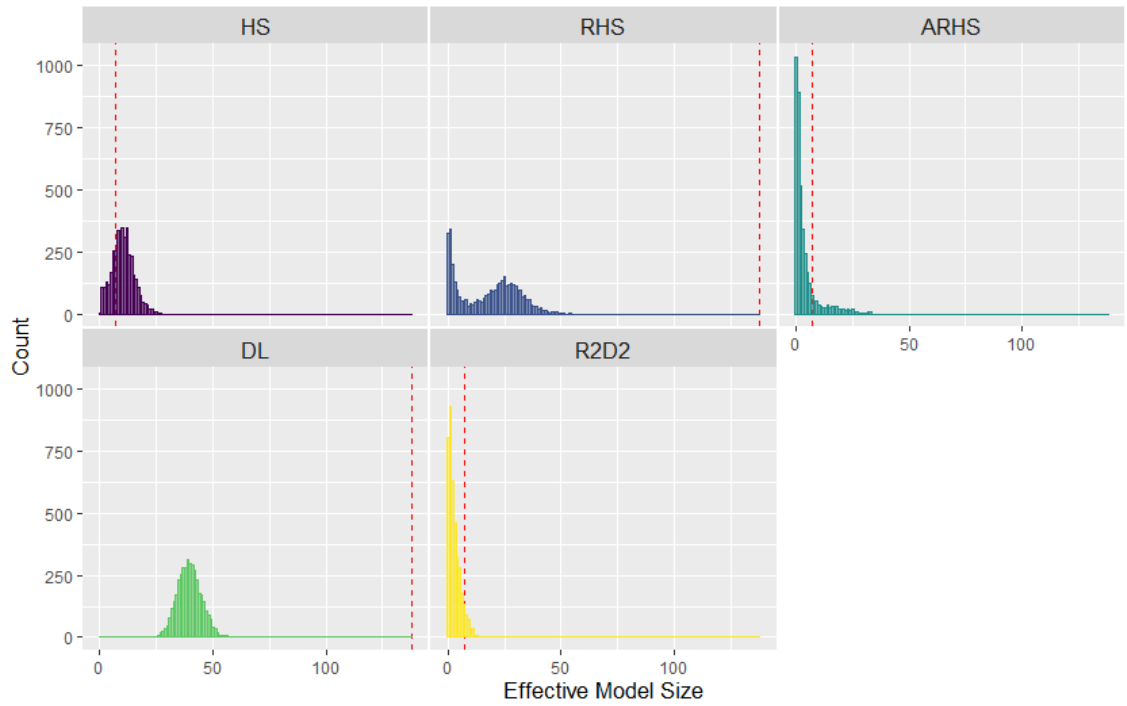


Figure E.10: Micro 2: SAVS posterior model sizes for the HS, RHS, A-RHS, DL, R2D2 priors and the posterior model size distributions for the SSVS and SSVS-t model. The red-dashed line indicated the model size chosen by Projpred

E.1.3 Posterior Inclusion Probabilities SAVS

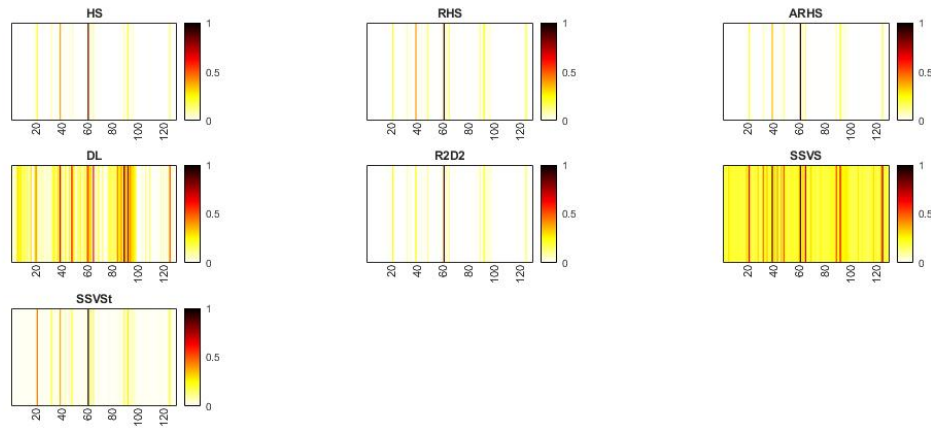


Figure E.11: Macro 1: Posterior inclusion probabilities model for the HS, RHS, A-RHS, DL, R2D2 priors based on SAVS and SSVS and SSVSt based on their posterior

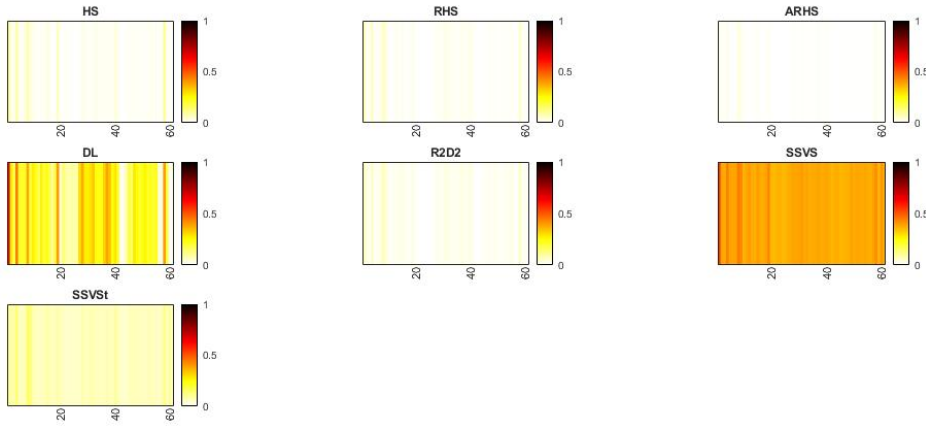


Figure E.12: Macro 2: Posterior inclusion probabilities model for the HS, RHS, A-RHS, DL, R2D2 priors based on SAVS and SSVS and SSVSt based on their posterior

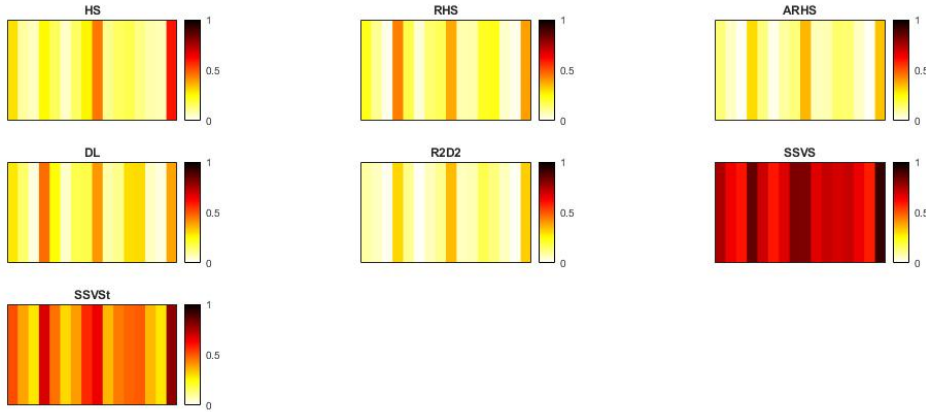


Figure E.13: Finance 1: Posterior inclusion probabilities model for the HS, RHS, A-RHS, DL, R2D2 priors based on SAVS and SSVS and SSVSt based on their posterior

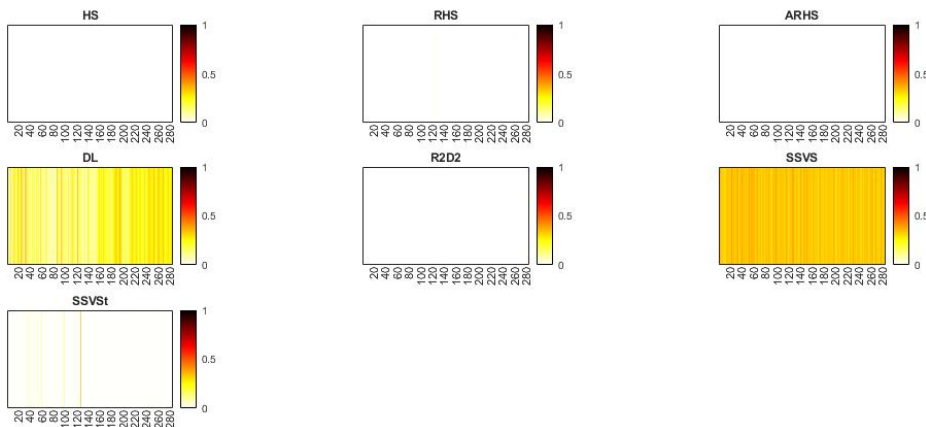


Figure E.14: Micro 1: Posterior inclusion probabilities model for the HS, RHS, A-RHS, DL, R2D2 priors based on SAVS and SSVS and SSVSt based on their posterior

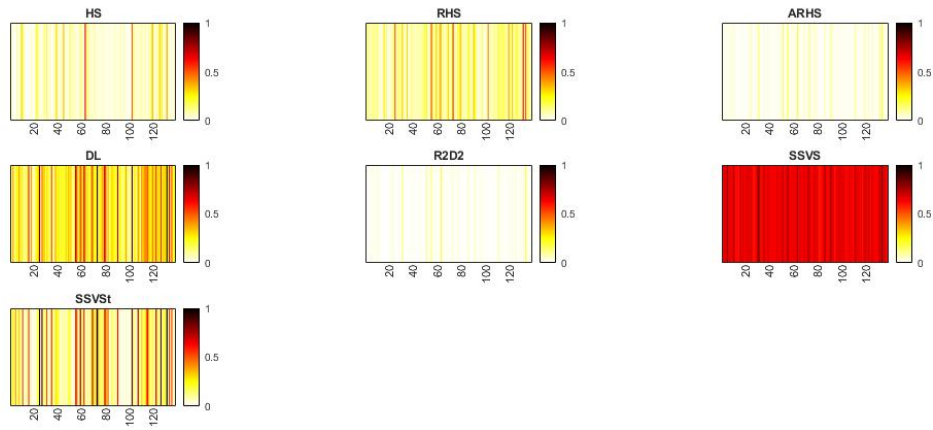


Figure E.15: Micro 2: Posterior inclusion probabilities model for the HS, RHS, A-RHS, DL, R2D2 priors based on SAVS and SSVs and SSVst based on their posterior

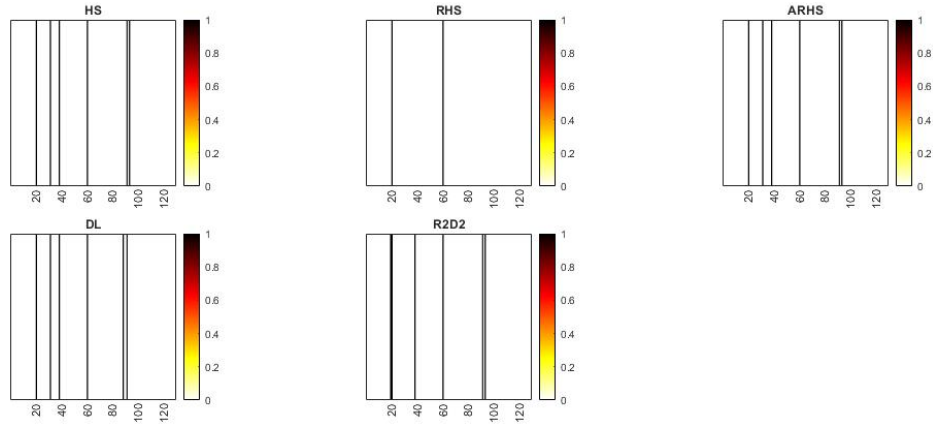


Figure E.16: Macro 1: The black line indicates the variables chosen by Projpred

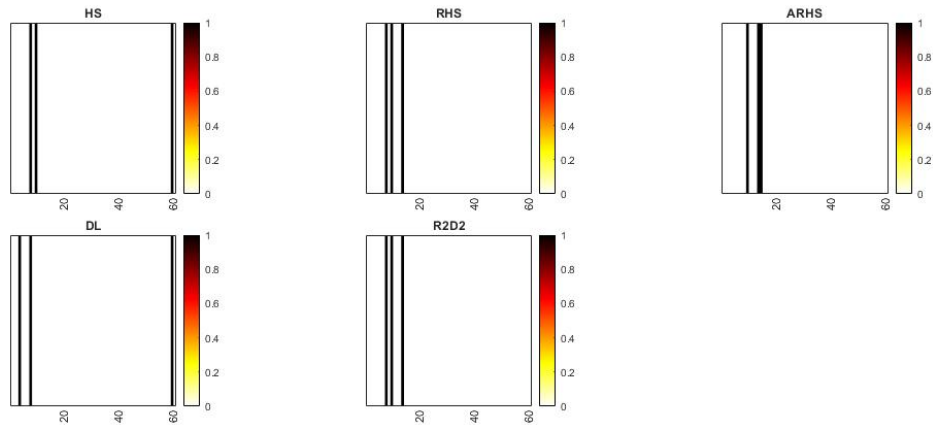


Figure E.17: Macro 2: The black line indicates the variables chosen by Projpred

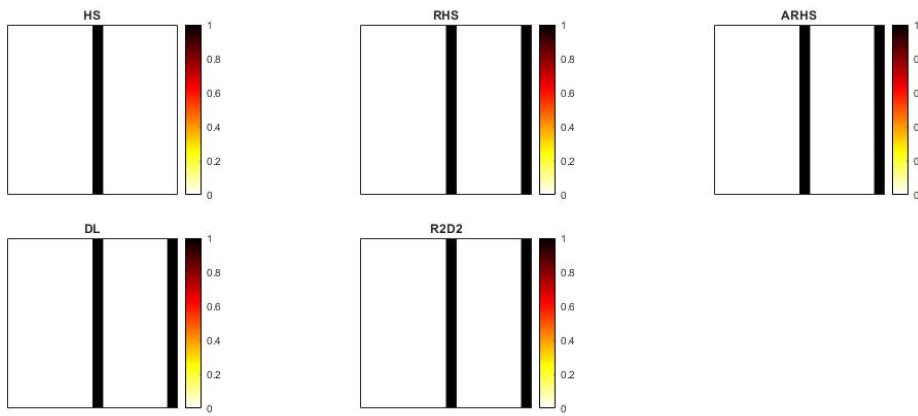


Figure E.18: Finance 1: The black line indicates the variables chosen by Projpred

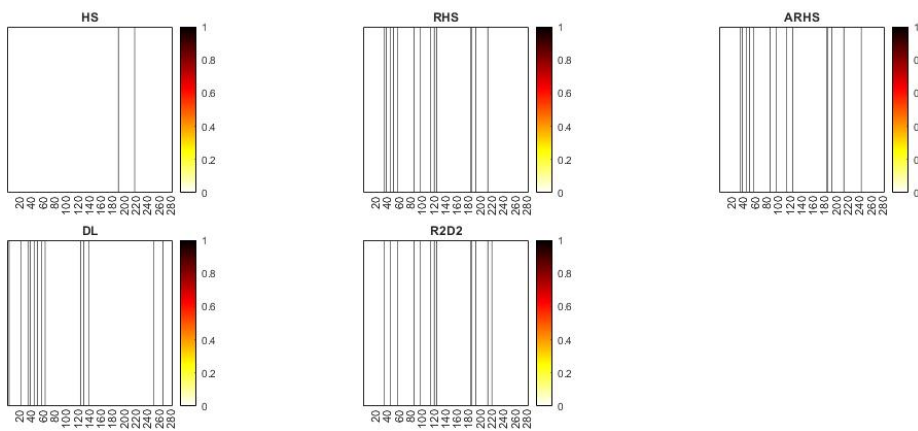


Figure E.19: Micro 1: The black line indicates the variables chosen by Projpred

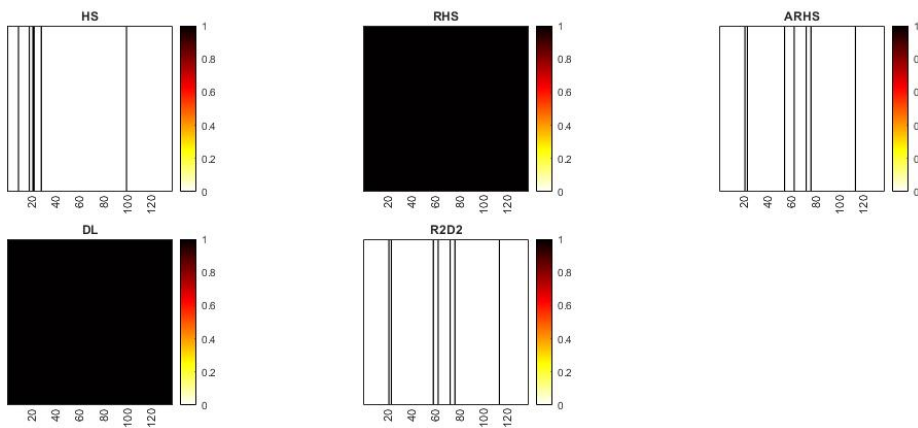


Figure E.20: Micro 2: The black line indicates the variables chosen by Projpred

E.1.4 ELPD Graphs

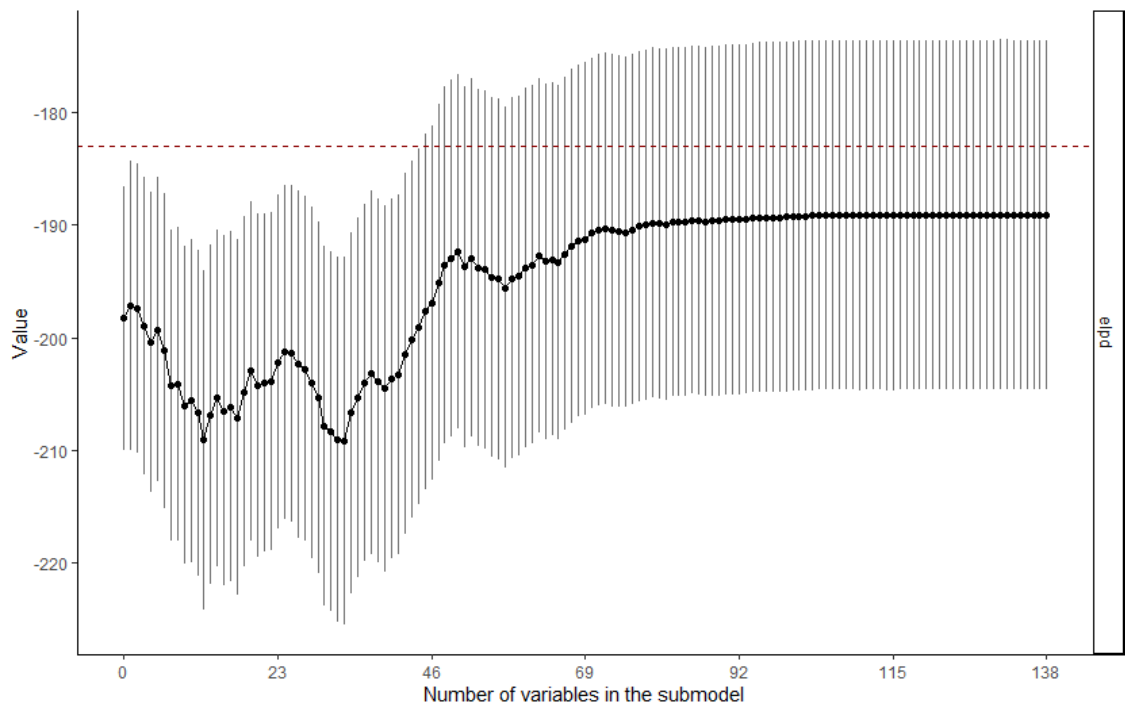


Figure E.21: ELPD scores for model sizes, RHS, Micro 2.

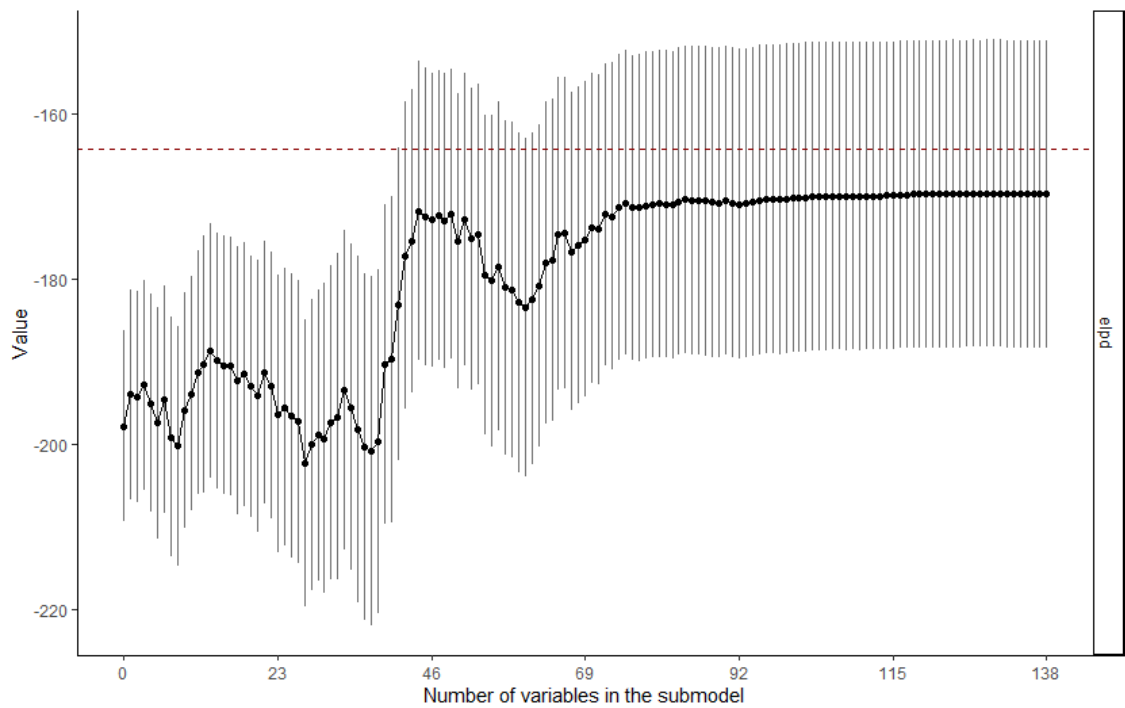


Figure E.22: ELPD scores for model sizes, DL, Micro 2.

Appendix F

Published Version of Chapter 2

F.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 chapter 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.

¹ The 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.

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 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 2.3, 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 2.4, we present results based on our empirical application of nowcasting U.S. GDP growth, which is followed in section 2.5 by the results from our simulation study. Finally, Section 2.6 concludes with a discussion and avenues for future research.

F.2 Bayesian Structural Time Series Model

F.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:

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

(F.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 ($\mathcal{N}\text{-}\mathcal{G}^{-1}$) 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.

F.2.2 Model enhancements

Non-Centred Bayesian Structural Time Series

In line with previous nowcasting studies, this chapter focuses on nowcasting GDP *growth* rather than levels. However, two problems arise when applying model (F.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 \mathcal{G}^{-1} priors with no prior mass on zero, as implemented for Bayesian linear state space methods, can bias posterior state variances away from zero, thereby potential leading to false support for state dynamics which can hurt forecast performance.

We extend model (F.1) to flexibly let the data shut down state dynamics, and therefore broaden the applicability of model (F.1), by adopting the non-centred

² Modelling 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 (2012a).

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 F.4. Testing for zero posterior variance would be very challenging in a frequentist hypothesis testing approach because the null hypothesis of constant state in model (F.1) lies on the boundary of the parameter space.

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

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

and

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

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 (F.2) and (F.3) is equivalent to (F.1), let:

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

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

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

which recovers (F.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, 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

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

⁴ For further discussion, please see section F.3

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:

$$\theta \sim N(\theta_0, V_\theta). \quad (\text{F.6})$$

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 Jeliaskov (2009) which is outlined in Appendix (A.1.4) along with the state posteriors. In contrast to FFBS type algorithms, it samples the states without recursive estimation which speeds up computation significantly.

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:

$$\begin{aligned} \beta_j | \gamma_j, \delta_j^2 &\sim \gamma_j N(0, \delta_j^2) + (1 - \gamma_j) N(0, c \times \delta_j^2) \\ \delta_j^2 &\sim \mathcal{G}(a_1, a_2) \\ \gamma_j &\sim \text{Ber}(\pi_0) \\ \pi_0 &\sim \mathcal{B}(b_1, b_2) \end{aligned} \quad (\text{F.7})$$

where $j \in (1, \dots, K)$ and \mathcal{G} , Ber and \mathcal{B} denote the Gamma, Bernoulli and Beta distribution respectively. 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 through 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, (F.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.1.1). 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.

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:

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

The idea of this family of priors is that the global scale ϑ controls the overall shrinkage applied to the regression, while the local scale φ_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 ϑ and φ_j :

$$\begin{aligned} \varphi_j &\sim C_+(0, 1) \\ \vartheta &\sim C_+(0, 1) \end{aligned} \tag{F.9}$$

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 (Piiironen et al., 2017) (see section A.1.2 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 (??).

Nevertheless, fitting the horseshoe prior can be challenging when the scale pa-

rameters 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.1.2) 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.1.1).

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:

$$\psi_j = \text{sign}(\hat{\beta}_j) \|X_j\|^{-2} \max(|\hat{\beta}_j| \|X_j\| - \zeta_j, 0), \quad (\text{F.10})$$

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 ζ_j in (6.17) 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

$$\zeta_j = \frac{1}{|\beta_j|^2}, \quad (\text{F.11})$$

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 similarity to lasso style regularisation, the solution to (6.17) can be obtained by the following minimisation problem which is closely related to the adaptive lasso (Zou, 2006):

$$\bar{\psi} = \underset{\psi}{\text{argmin}} \left\{ \frac{1}{2} \|X\hat{\beta} - X\psi\|^2 + \sum_{j=1}^K \zeta_j |\psi_j| \right\}. \quad (\text{F.12})$$

Here, $\bar{\psi}$ 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).

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

Sampling Algorithm

With the conditional posteriors for the regression and state components at hand (see Appendix A.1.1), 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 F.2.2, 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.1.4. As discussed in Section F.2.2, 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 (F.10) on an iteration basis.

F.3 Data

F.3.1 Mixing Frequencies

In this chapter, 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 (Bańbura 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 (Forni et al., 2015; Forni and Marcellino, 2014), as is the case in our application. Switching notation from equation (F.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 β . This vertical realignment is visualised for a single representative regressor below:

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

F.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 (3.1) 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

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

⁷ We 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 chapter remain qualitatively identical. Details are available upon request.

⁸ The 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 evaluation.

typically during the last week of a given month. Real-time vintages are downloaded from the FRED database using the ‘FredFetch’ Matlab package.⁹

F.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 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

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

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

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

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
- **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 A.13 in appendix). Topics around housing have been shown to be indicative of housing prices (Wu and Brynjolfsson, 2015; Askitas 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).

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

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., 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 F.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

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

¹⁴ To 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.

¹⁵ Due 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.

sampled.

F.3.4 Understanding Google Trends

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

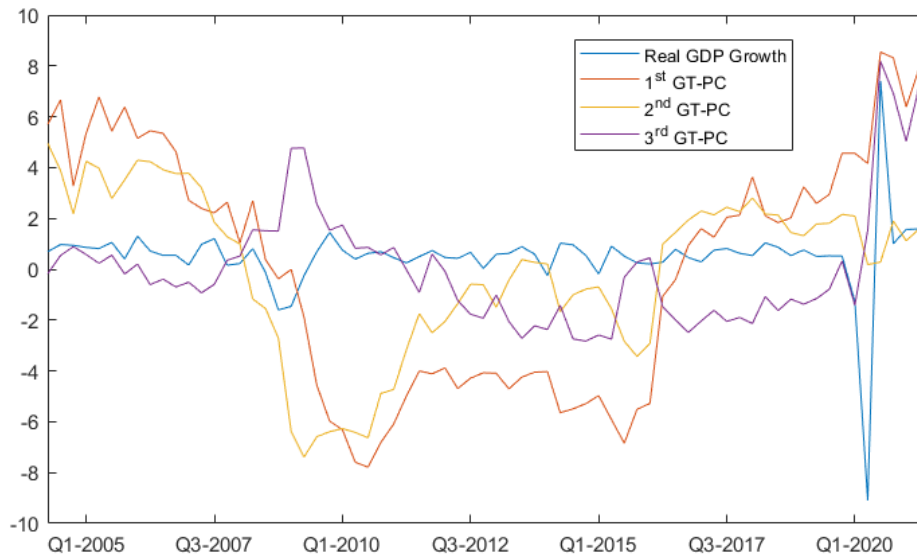


Figure F.1: First 3 principal components of the GT data set and 2nd vintage real GDP growth.

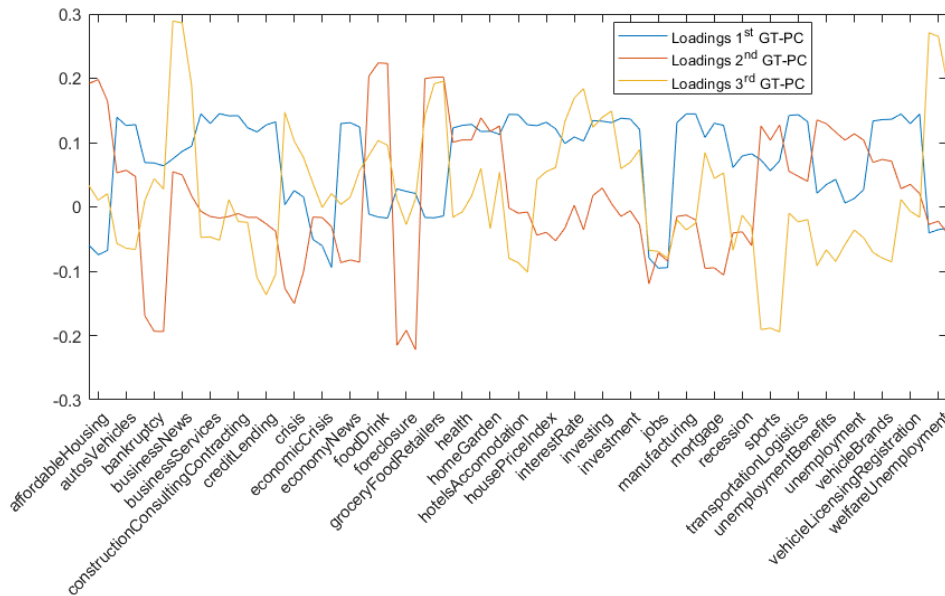


Figure F.2: Loadings for first 3 principal components of the GT data set.

The first three principal components show very heterogeneous behaviour, but the dynamics conform to the economic intuition suggested from the loadings. The

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	USEPUINDEXM
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	USEPUINDEXM
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	USEPUINDEXM
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 F.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.

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 A.13 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 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.

F.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 (F.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 UMPI 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 UMPI index and negatively with the BAA spread (which increases during deteriorating financial conditions), indicating that it indeed captures something close to consumer 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.

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

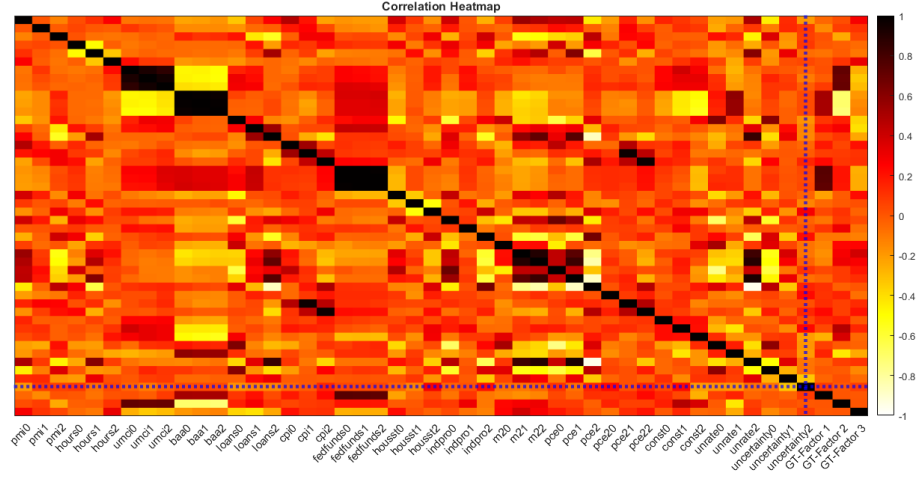


Figure F.3: Heatmap of sample correlation matrix.

F.4 Nowcasting U.S. Real GDP Growth

The predictive model used to generate in-sample and out-of-sample predictions is given in equations (F.2)-(F.3) where the first $T = 45$ observations are used as training sample.¹⁷ We estimate three variants of the model based on priors (F.7, F.9, and F.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 periods within the real-time calendar (Table F.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:

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

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

¹⁷ As 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.

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):

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

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

where, for brevity of notation, $\Upsilon_{1:T+j-1}$ 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 (3.11), $y_{T+j}^{v,A,B}$ are independently drawn from the posterior predictive density $p(y_{T+j}^v | \Omega_{T+j-1}^v | y_T)$.

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 (F.14, F.15, and F.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.1.4.

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

¹⁹ Although 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.

²⁰ The 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.

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

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

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.

F.4.1 In-Sample Results

Figure F.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 F.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 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.

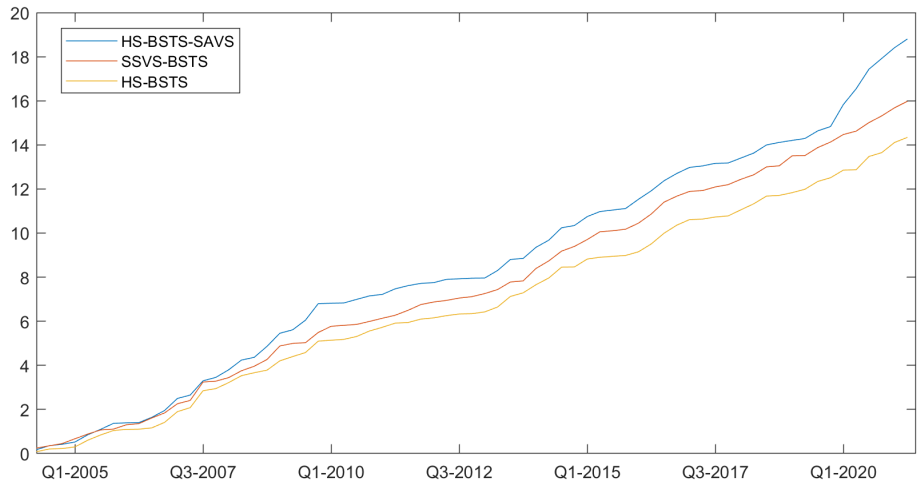


Figure F.4: Cumulative one-step-ahead forecast errors in-sample from 3 different models: (1) SSVS-BSTS, (2) HS-BSTS and (3) HS-BSTS-SAVS

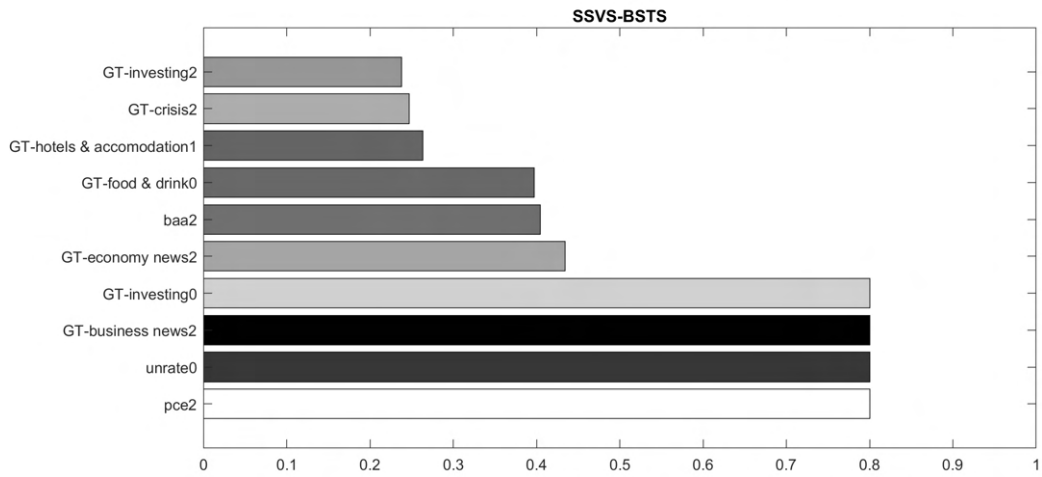


Figure F.5: Posterior inclusion probabilities of the SSVS-BSTS model.

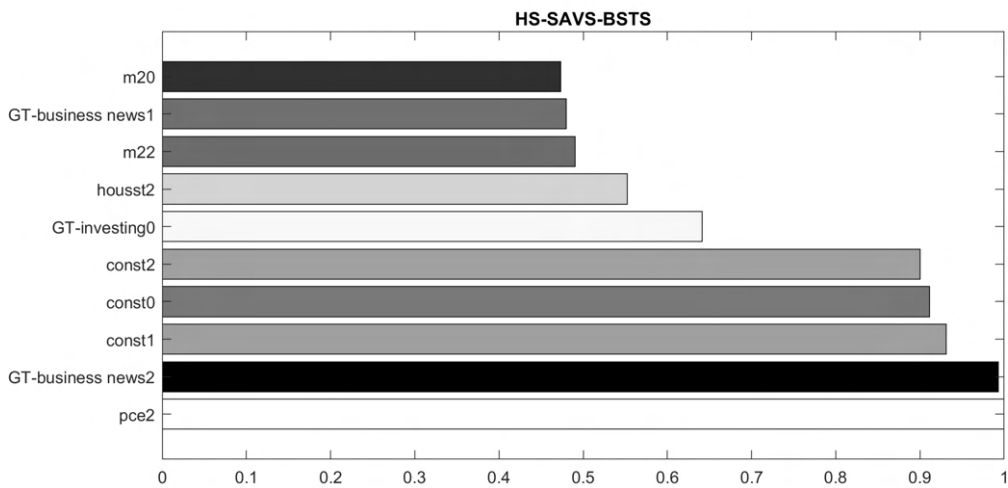


Figure F.6: Posterior inclusion probabilities of the HS-SAVS-BSTS model.

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 F.5 and F.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 A.11), 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 A.12) 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., 2020; 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 A.13, 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 F.5 and F.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

²¹ It 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.

highest marginal likelihood. Continuous shrinkage priors, on the other hand, make use of all covariates since they are always included in the model.

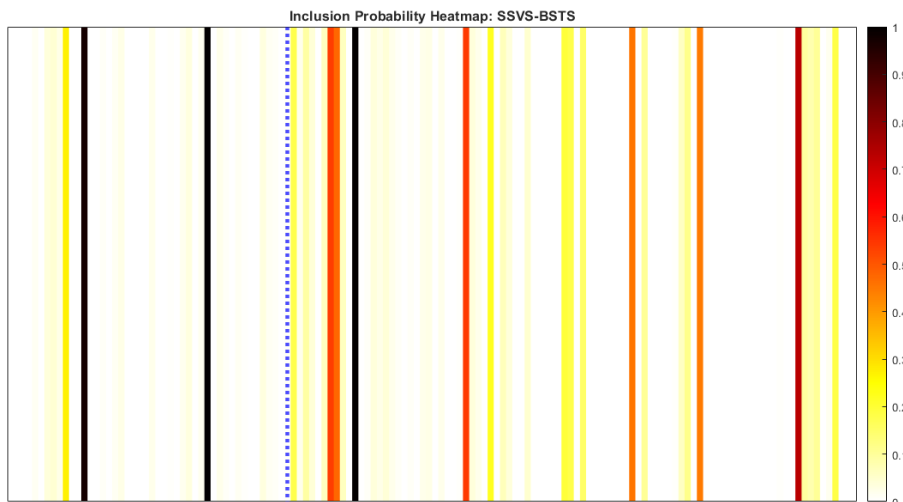


Figure F.7: Posterior Inclusion probability heatmaps for the SSVS-BSTS. Inclusion probabilities to the left of the dashed line pertain to the macro data set, and to the right, the Google search data.

In line with this discussion, the posterior inclusion probability heatmaps in Figures F.7 and F.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 and Google Trends) and within mixed frequency information. And secondly, different sparsity patterns can emerge depending on the prior used. Section F.5 further investigates robustness of the proposed priors to different sparsity settings, and provide recommendations.

Finally, our in-sample results clearly demonstrate (Figure F.9) that there is support in the data for a local trend, but not a local linear trend: the posterior for σ^τ

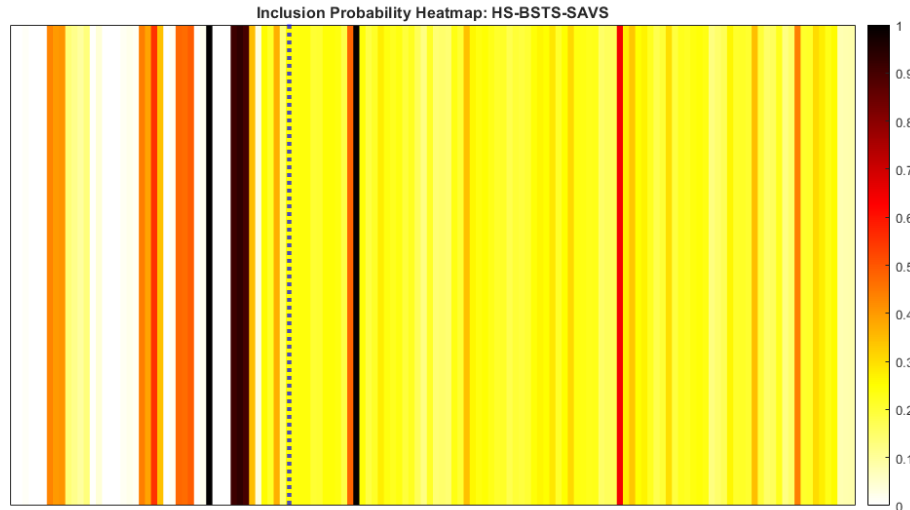


Figure F.8: Posterior Inclusion probability heatmaps for the HS-SAVS-BSTS. Inclusion probabilities to the left of the dashed line pertain to the macro data set, and to the right, the Google search data.

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.

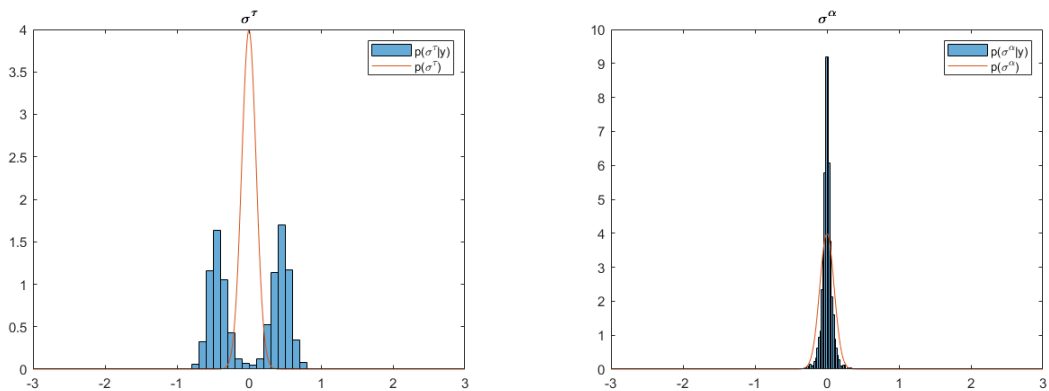


Figure F.9: Distribution of the (left) trend state standard deviation and (right) slope standard deviation for the HS-BSTS model.

F.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 F.3. 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 F.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 F.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 consequently poor nowcasting performance.

²² The 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.

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

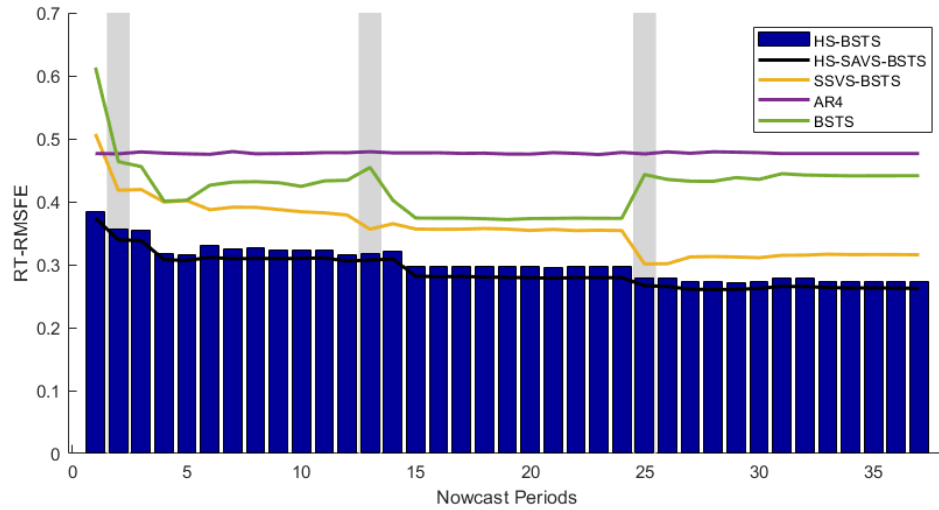


Figure F.10: Real-Time RMSFE of all competing models. The RT-RMSFE for the BSTS are plotted on the right axis. Grey vertical bars indicate nowcast periods in which GT are published.

Similar to the real-time point forecasts, we plot real-time LPDS (RT-LPDS) and CRPS (RT-CRPS) in Figures F.11 and F.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.

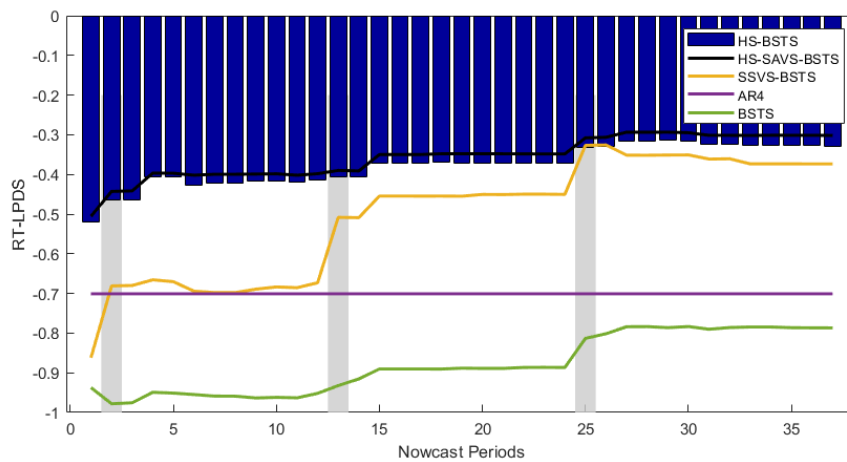


Figure F.11: Real-Time log-predictive density scores (RT-LPDS) for all competing models. The RT-LPDS for the BSTS are plotted on the right axis. Grey vertical bars indicate nowcast periods in which GT are published.

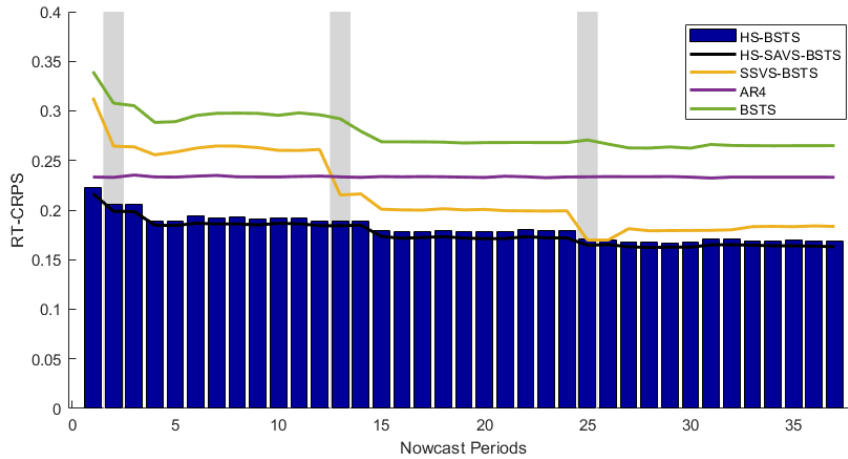


Figure F.12: Real-Time CRPS scores (RT-CRPS) for all competing models. The RT-CRPS for the BSTS are plotted on the right axis. Grey vertical bars indicate nowcast periods in which GT are published.

F.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 F.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

²⁴ Higher 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.

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

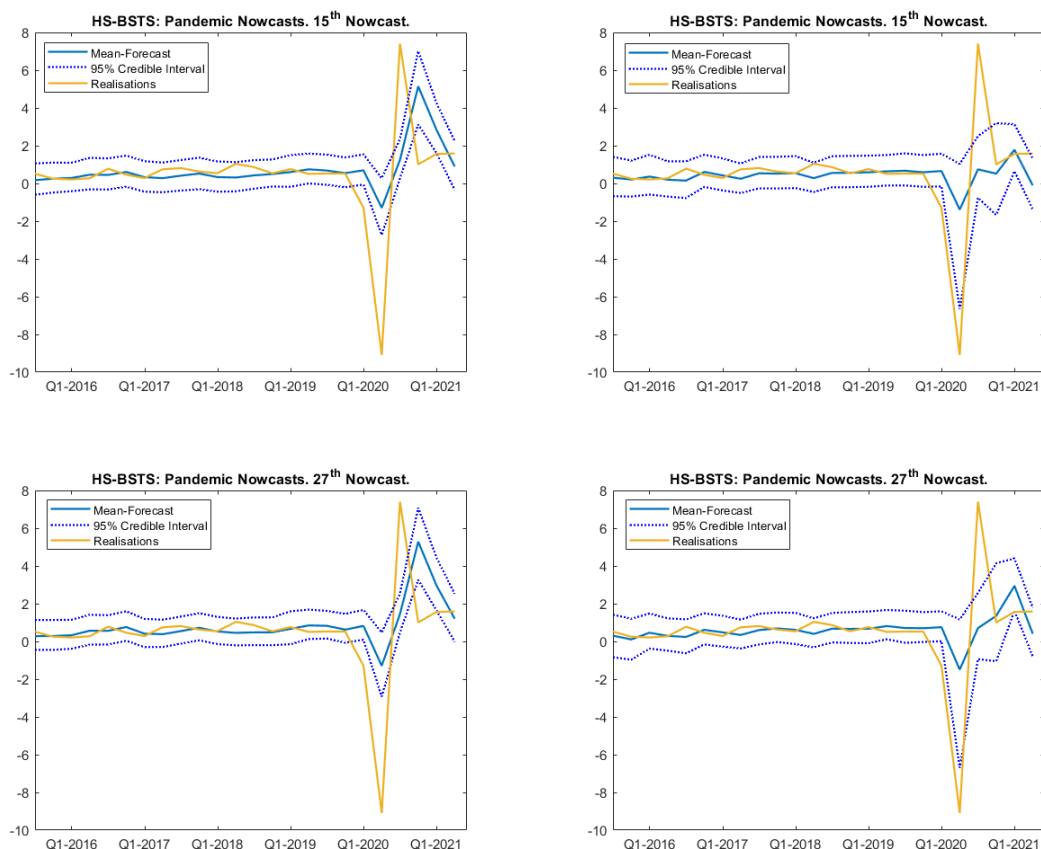


Figure F.13: Predictive distributions for the macro only data set (left column) and the macro + Google Trends data set (right columns) for the 15th (upper row) and 27th (lower row) nowcast period.

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.

F.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 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

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

down-weighting outliers through the error covariances.

The estimated model uses the same regression and state components as model (F.2)-(F.3). However, we assume $\epsilon_t \sim N(0, \sigma^2 \lambda_t)$, where λ_t is distributed as $\mathcal{G}^{-1}(\nu/2, \nu/2)$ where \mathcal{G}^{-1} denotes the inverse-Gamma distribution and ν the degree of freedom parameter of the t-distribution. 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.1.5). Treating ν as a random variable, figure (A.9), 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.

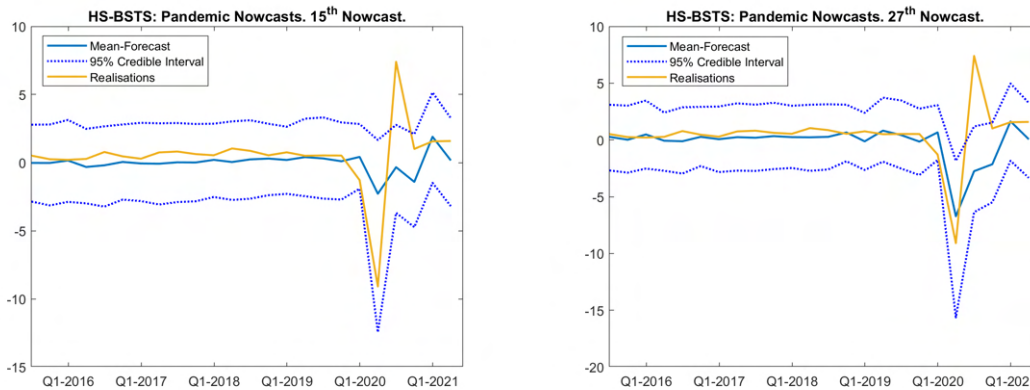


Figure F.14: Predictive distributions for the full HS-BSTS-t model.

The nowcasts from this model (Figure F.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 (A.10) 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 (A.8) 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.

F.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 chapter 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 (F.2)-(F.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

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

and the dense coefficient vector is

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

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 (F.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

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

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.

$(\sigma^\tau, \sigma^\alpha)$	Sparse				Dense			
	(0.5,0)	(0,0.5)	(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 F.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 F.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

²⁷ Note that we do not report DS tests for the original BSTS model. This is due to the fact that the prior on the state variance has no mass on zero and therefore is not testable.

F.6 Conclusion

In this chapter, 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 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.

Bibliography

- Adams, P., T. Adrian, N. Boyarchenko, and D. Giannone (2020). Forecasting macroeconomic risks.
- Adrian, T., N. Boyarchenko, and D. Giannone (2019). Vulnerable growth. *American Economic Review* 109(4), 1263–89.
- Afrabandpey, H., T. Peltola, J. Piironen, A. Vehtari, and S. Kaski (2019). Making bayesian predictive models interpretable: a decision theoretic approach. *arXiv preprint arXiv:1910.09358*.
- Alexopoulos, M. and J. Cohen (2015). The power of print: Uncertainty shocks, markets, and the economy. *International Review of Economics & Finance* 40, 8–28.
- Alhamzawi, R. and K. Yu (2013). Conjugate priors and variable selection for bayesian quantile regression. *Computational Statistics & Data Analysis* 64, 209–219.
- Alhamzawi, R., K. Yu, and D. F. Benoit (2012). Bayesian adaptive lasso quantile regression. *Statistical Modelling* 12(3), 279–297.
- Almon, S. (1965). The distributed lag between capital appropriations and expenditures. *Econometrica: Journal of the Econometric Society*, 178–196.
- Andreou, E., E. Ghysels, and A. Kourtellis (2010). Regression models with mixed sampling frequencies. *Journal of Econometrics* 158(2), 246–261.
- Anesti, N., A. B. Galvão, and S. Miranda-Agrippino (2018). Uncertain kingdom: nowcasting gdp and its revisions.
- Anesti, N., S. Hayes, A. Moreira, and J. Tasker (2017). Peering into the present: the bank’s approach to gdp nowcasting. *Bank of England Quarterly Bulletin*, Q2.
- Antolin-Diaz, J., T. Drechsel, and I. Petrella (2017). Tracking the slowdown in long-run gdp growth. *Review of Economics and Statistics* 99(2), 343–356.
- Antolin-Diaz, J., T. Drechsel, and I. Petrella (2021). Advances in nowcasting economic activity: Secular trends, large shocks and new data.

- 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).
- Armagan, A., D. B. Dunson, and J. Lee (2013). Generalized double pareto shrinkage. *Statistica Sinica* 23(1), 119.
- Askitas, N. and K. F. Zimmermann (2009). Google econometrics and unemployment forecasting.
- Azzalini, A. and A. Capitanio (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 65(2), 367–389.
- Bai, J. and S. Ng (2008). Forecasting economic time series using targeted predictors. *Journal of Econometrics* 146(2), 304–317.
- Baker, S. R., N. Bloom, and S. J. Davis (2016). Measuring economic policy uncertainty. *The quarterly journal of economics* 131(4), 1593–1636.
- Bañbura, 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.
- Bañbura, M. and M. Modugno (2014). Maximum likelihood estimation of factor models on datasets with arbitrary pattern of missing data. *Journal of Applied Econometrics* 29(1), 133–160.
- Banerjee, A. and M. Marcellino (2006). Are there any reliable leading indicators for us inflation and gdp growth? *International Journal of Forecasting* 22(1), 137–151.
- Barbieri, M. M. and J. O. Berger (2004). Optimal predictive model selection. *The annals of statistics* 32(3), 870–897.
- Barbieri, M. M., J. O. Berger, E. I. George, and V. Ročková (2021). The median probability model and correlated variables. *Bayesian Analysis* 16(4), 1085–1112.
- Barro, R. J. (1991). Economic growth in a cross section of countries. *The quarterly journal of economics* 106(2), 407–443.
- Baumeister, C., D. Leiva-León, and E. R. Sims (2021). Tracking weekly state-level economic conditions. Technical report, National Bureau of Economic Research.
- Belloni, A., D. Chen, V. Chernozhukov, and C. Hansen (2012). Sparse models and methods for optimal instruments with an application to eminent domain. *Econometrica* 80(6), 2369–2429.

- 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, and B. Willard (2017). The horseshoe+ estimator of ultra-sparse signals. *Bayesian Analysis* 12(4), 1105–1131.
- 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.
- Bhattacharya, A., D. Pati, N. S. Pillai, and D. B. Dunson (2015). Dirichlet–laplace priors for optimal shrinkage. *Journal of the American Statistical Association* 110(512), 1479–1490.
- Billio, M., R. Casarin, F. Ravazzolo, and H. K. Van Dijk (2013). Time-varying combinations of predictive densities using nonlinear filtering. *Journal of Econometrics* 177(2), 213–232.
- Bitto, A. and S. Frühwirth-Schnatter (2019). Achieving shrinkage in a time-varying parameter model framework. *Journal of Econometrics* 210(1), 75–97.
- Board, F. S. et al. (2011). Macroprudential policy tools and frameworks: Progress report to g20. *Basel: Bank for International Settlements*.
- 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.
- Boss, J., J. Datta, X. Wang, S. K. Park, J. Kang, and B. Mukherjee (2021). Group inverse-gamma gamma shrinkage for sparse regression with block-correlated predictors. *arXiv preprint arXiv:2102.10670*.
- Breheeny, P. and J. Huang (2015). Group descent algorithms for nonconvex penalized linear and logistic regression models with grouped predictors. *Statistics and computing* 25(2), 173–187.
- Cadonna, A., S. Frühwirth-Schnatter, and P. Knaus (2020). Triple the gamma—a unifying shrinkage prior for variance and variable selection in sparse state space and tvp models. *Econometrics* 8(2), 20.
- Carpenter, B., A. Gelman, M. D. Hoffman, D. Lee, B. Goodrich, M. Betancourt, M. Brubaker, J. Guo, P. Li, and A. Riddell (2017). Stan: A probabilistic programming language. *Journal of statistical software* 76(1), 1–32.

- Carrière-Swallow, Y. and F. Labbé (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, and M. Marcellino (2020a). Assessing international commonality in macroeconomic uncertainty and its effects. *Journal of Applied Econometrics* 35(3), 273–293.
- Carriero, A., T. E. Clark, and M. G. Marcellino (2020b). Nowcasting tail risks to economic activity with many indicators.
- Carriero, A., T. E. Clark, M. G. Marcellino, and E. Mertens (2021). Addressing covid-19 outliers in bvars with stochastic volatility.
- Carriero, A., A. B. Galvão, and G. Kapetanios (2019). A comprehensive evaluation of macroeconomic forecasting methods. *International Journal of Forecasting* 35(4), 1226–1239.
- Carter, C. K. and R. Kohn (1994). On gibbs sampling for state space models. *Biometrika* 81(3), 541–553.
- Carvalho, C. M., N. G. Polson, and J. G. Scott (2010). The horseshoe estimator for sparse signals. *Biometrika* 97(2), 465–480.
- Casella, G., M. Ghosh, J. Gill, and M. Kyung (2010). Penalized regression, standard errors, and bayesian lassos. *Bayesian analysis* 5(2), 369–411.
- 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 macroeconometrics. *Manuscript available at <http://joshuachan.org>*.
- Chan, J. C. (2017b). The stochastic volatility in mean model with time-varying parameters: An application to inflation modeling. *Journal of Business & Economic Statistics* 35(1), 17–28.
- Chan, J. C., E. Eisenstat, and R. W. Strachan (2020). Reducing the state space dimension in a large tvp-var. *Journal of Econometrics* 218(1), 105–118.

- 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., D. B. Dunson, C. Reed, and K. Yu (2013). Bayesian variable selection in quantile regression. *Statistics and its Interface* 6(2), 261–274.
- 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.
- Chernozhukov, V., I. Fernández-Val, and A. Galichon (2010). Quantile and probability curves without crossing. *Econometrica* 78(3), 1093–1125.
- Chernozhukov, V., C. Hansen, Y. Liao, et al. (2017). A lava attack on the recovery of sums of dense and sparse signals. *The Annals of Statistics* 45(1), 39–76.
- Chipman, H. A., E. I. George, and R. E. McCulloch (2010). Bart: Bayesian additive regression trees. *The Annals of Applied Statistics* 4(1), 266–298.
- 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.
- Clements, M. P. and A. B. Galvão (2009). Forecasting us output growth using leading indicators: An appraisal using midas models. *Journal of Applied Econometrics* 24(7), 1187–1206.
- Clyde, M. and E. I. George (2004). Model uncertainty. *Statistical science*, 81–94.
- 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.
- Coulombe, P. G., M. Leroux, D. Stevanovic, and S. Surprenant (2020). How is machine learning useful for macroeconomic forecasting? *arXiv preprint arXiv:2008.12477*.
- Cross, J. L., C. Hou, G. Koop, and A. Poon (2021). Macroeconomic forecasting with large stochastic volatility in mean vars.
- 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.
- Croushore, D. and T. Stark (2001). A real-time data set for macroeconomists. *Journal of econometrics 105*(1), 111–130.
- De Mol, C., D. Giannone, and L. Reichlin (2008). Forecasting using a large number of predictors: Is bayesian shrinkage a valid alternative to principal components? *Journal of Econometrics 146*(2), 318–328.
- Devroye, L. (2014). Random variate generation for the generalized inverse gaussian distribution. *Statistics and Computing 24*(2), 239–246.
- Diebold, F. X., T. A. Gunther, and A. S. Tay (1998). Evaluating density forecasts with applications to financial risk management. *International Economic Review 39*(4), 863–883.
- Diebold, F. X. and R. S. Mariano (2002). Comparing predictive accuracy. *Journal of Business & economic statistics 20*(1), 134–144.
- Doan, T., R. Litterman, and C. Sims (1984). Forecasting and conditional projection using realistic prior distributions. *Econometric reviews 3*(1), 1–100.
- Donohue III, J. J. and S. D. Levitt (2001). The impact of legalized abortion on crime. *The Quarterly Journal of Economics 116*(2), 379–420.
- Dupuis, J. A. and C. P. Robert (2003). Variable selection in qualitative models via an entropic explanatory power. *Journal of Statistical Planning and Inference 111*(1-2), 77–94.
- 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.
- Edge, R. M., M. T. Kiley, and J.-P. Laforte (2010). A comparison of forecast performance between federal reserve staff forecasts, simple reduced-form models, and a dsge model. *Journal of Applied Econometrics 25*(4), 720–754.
- Engle, R. F. and S. Manganelli (2004). Caviar: Conditional autoregressive value at risk by regression quantiles. *Journal of Business & Economic Statistics 22*(4), 367–381.
- 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.

- Fan, J. and R. Li (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American statistical Association* 96(456), 1348–1360.
- Fava, B. and H. F. Lopes (2021). The illusion of the illusion of sparsity: An exercise in prior sensitivity. *Brazilian Journal of Probability and Statistics* 35(4), 699–720.
- Ferrara, L., M. Mogliani, and J.-G. Sahuc (2020). High-frequency monitoring of growth-at-risk.
- 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.
- Figueres, J. M. and M. Jarociński (2020). Vulnerable growth in the euro area: Measuring the financial conditions. *Economics Letters*, 109126.
- 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.
- Froni, 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.
- Froni, C., M. Marcellino, and C. Schumacher (2015). Unrestricted mixed data sampling (midas): Midas regressions with unrestricted lag polynomials. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 178(1), 57–82.
- Friedman, J., T. Hastie, H. Höfling, and R. Tibshirani (2007). Pathwise coordinate optimization. *The annals of applied statistics* 1(2), 302–332.
- Friedman, J., T. Hastie, and R. Tibshirani (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of statistical software* 33(1), 1.
- Friedman, J., T. Hastie, R. Tibshirani, et al. (2001). *The elements of statistical learning*, Volume 1. Springer series in statistics New York.
- 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. F. Lopes (2018). Sparse bayesian factor analysis when the number of factors is unknown. *arXiv preprint arXiv:1804.04231*.
- Frühwirth-Schnatter, S. and H. Wagner (2010). Stochastic model specification search for gaussian and partial non-gaussian state space models. *Journal of Econometrics* 154(1), 85–100.
- Gaglianone, W. P. and L. R. Lima (2012). Constructing density forecasts from quantile regressions. *Journal of Money, Credit and Banking* 44(8), 1589–1607.
- Gaglianone, W. P., L. R. Lima, O. Linton, and D. R. Smith (2011). Evaluating value-at-risk models via quantile regression. *Journal of Business & Economic Statistics* 29(1), 150–160.
- Gefang, D., G. Koop, and A. Poon (2020). Computationally efficient inference in large bayesian mixed frequency vars. *Economics Letters* 191, 109120.
- Gelman, A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin (2013). *Bayesian data analysis*. CRC press.
- Gelman, A., A. Vehtari, D. Simpson, C. C. Margossian, B. Carpenter, Y. Yao, L. Kennedy, J. Gabry, P.-C. Bürkner, and M. Modrák (2020). Bayesian workflow. *arXiv preprint arXiv:2011.01808*.
- 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.
- George, E. I. and R. E. McCulloch (1997). Approaches for bayesian variable selection. *Statistica sinica*, 339–373.
- George, E. I., D. Sun, and S. Ni (2008). Bayesian stochastic search for var model restrictions. *Journal of Econometrics* 142(1), 553–580.
- 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.
- Ghysels, E. and M. Marcellino (2018). *Applied economic forecasting using time series methods*. Oxford University Press.

- Ghysels, E., A. Sinko, and R. Valkanov (2007). Midas regressions: Further results and new directions. *Econometric reviews* 26(1), 53–90.
- Giannone, D., M. Lenza, and G. E. Primiceri (2019). Priors for the long run. *Journal of the American Statistical Association* 114(526), 565–580.
- Giannone, D., M. Lenza, and G. E. Primiceri (2021). Economic predictions with big data: The illusion of sparsity. *Econometrica* 89(5), 2409–2437.
- Giannone, D., F. Monti, and L. Reichlin (2016). Exploiting the monthly data flow in structural forecasting. *Journal of Monetary Economics* 84, 201–215.
- Giannone, D., L. Reichlin, and D. Small (2008). Nowcasting: The real-time informational content of macroeconomic data. *Journal of monetary economics* 55(4), 665–676.
- Giglio, S., B. Kelly, and S. Pruitt (2016). Systemic risk and the macroeconomy: An empirical evaluation. *Journal of Financial Economics* 119(3), 457–471.
- Gneiting, T. and A. E. Raftery (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American statistical Association* 102(477), 359–378.
- Gneiting, T. and R. Ranjan (2011). Comparing density forecasts using threshold-and quantile-weighted scoring rules. *Journal of Business & Economic Statistics* 29(3), 411–422.
- Goodfriend, M. and R. G. King (2005). The incredible volcker disinflation. *Journal of Monetary Economics* 52(5), 981–1015.
- Goutis, C. and C. P. Robert (1998). Model choice in generalised linear models: A bayesian approach via kullback-leibler projections. *Biometrika* 85(1), 29–37.
- Gramacy, R. B. (2020). *Surrogates: Gaussian process modeling, design, and optimization for the applied sciences*. Chapman and Hall/CRC.
- 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.
- Griffin, J. E., P. J. Brown, et al. (2010). Inference with normal-gamma prior distributions in regression problems. *Bayesian analysis* 5(1), 171–188.
- Guérin, P. and M. Marcellino (2013). Markov-switching midas models. *Journal of Business & Economic Statistics* 31(1), 45–56.
- 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.

- Hahn, P. R. and C. M. Carvalho (2015). Decoupling shrinkage and selection in bayesian linear models: a posterior summary perspective. *Journal of the American Statistical Association* 110(509), 435–448.
- Harvey, A. C. (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.
- Hasenzagl, T., L. Reichlin, and G. Ricco (2020). Financial variables as predictors of real growth vulnerability.
- Hastie, T., R. Tibshirani, J. H. Friedman, and J. H. Friedman (2009). *The elements of statistical learning: data mining, inference, and prediction*, Volume 2. Springer.
- Hauzenberger, N., F. Huber, and G. Koop (2020). Dynamic shrinkage priors for large time-varying parameter regressions using scalable markov chain monte carlo methods. *arXiv preprint arXiv:2005.03906*.
- He, X., J. Jurečková, R. Koenker, and S. Portnoy (1990). Tail behavior of regression estimators and their breakdown points. *Econometrica: Journal of the Econometric Society*, 1195–1214.
- Hodrick, R. J. and E. C. Prescott (1997). Postwar us business cycles: an empirical investigation. *Journal of Money, credit, and Banking*, 1–16.
- Hoerl, A. E. and R. W. Kennard (1970). Ridge regression: applications to nonorthogonal problems. *Technometrics* 12(1), 69–82.
- Hörmann, W. and J. Leydold (2014). Generating generalized inverse gaussian random variates. *Statistics and Computing* 24(4), 547–557.
- Huber, F., G. Koop, and L. Onorante (2019). Inducing sparsity and shrinkage in time-varying parameter models. *arXiv preprint arXiv:1905.10787*.
- Huber, F., G. Koop, L. Onorante, M. Pfarrhofer, and J. Schreiner (2020). Nowcasting in a pandemic using non-parametric mixed frequency vars. *Journal of Econometrics*.
- Hubrich, K. and R. J. Tetlow (2015). Financial stress and economic dynamics: The transmission of crises. *Journal of Monetary Economics* 70, 100–115.
- 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.

- Johnson, N. L., S. Kotz, and N. Balakrishnan (1995). *Continuous univariate distributions, volume 2*, Volume 289. John Wiley & Sons.
- Jordà, Ò., M. Schularick, and A. M. Taylor (2015). Leveraged bubbles. *Journal of Monetary Economics* 76, S1–S20.
- 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 (2020). Making text count: economic forecasting using newspaper text. *Journal of Applied Econometrics*.
- Kapetanios, G., F. Papailias, et al. (2022). Real time indicators during the covid-19 pandemic individual predictors & selection. Technical report, Economic Statistics Centre of Excellence (ESCoE).
- Khare, K. and J. P. Hobert (2012). Geometric ergodicity of the gibbs sampler for bayesian quantile regression. *Journal of Multivariate Analysis* 112, 108–116.
- Kim, C.-J. and C. R. Nelson (1999). Has the us economy become more stable? a bayesian approach based on a markov-switching model of the business cycle. *Review of Economics and Statistics* 81(4), 608–616.
- Kim, H. H. and N. R. Swanson (2014). Forecasting financial and macroeconomic variables using data reduction methods: New empirical evidence. *Journal of Econometrics* 178, 352–367.
- Kim, S., N. Shephard, and S. Chib (1998). Stochastic volatility: likelihood inference and comparison with arch models. *The review of economic studies* 65(3), 361–393.
- Koenker, R. (2005). *Quantile regression*. New York: Cambridge University Press.
- Koenker, R. and G. Bassett (1978). Regression quantiles. *Econometrica* 46, 33–50.
- Koenker, R., V. Chernozhukov, X. He, and L. Peng (2017). Handbook of quantile regression.
- Koenker, R. and J. A. Machado (1999). Goodness of fit and related inference processes for quantile regression. *Journal of the american statistical association* 94(448), 1296–1310.
- Kohns, D. and A. Bhattacharjee (2020). Developments on the bayesian structural time series model: Trending growth. *arXiv preprint arXiv:2011.00938*.

- Kohns, D. and A. Bhattacharjee (2022). Nowcasting growth using google trends data: A bayesian structural time series model. *International Journal of Forecasting*.
- Kohns, D. and T. Szendrei (2020). Horseshoe prior bayesian quantile regression. *arXiv preprint arXiv:2006.07655*.
- Konrath, S., T. Kneib, and L. Fahrmeir (2008). Bayesian regularisation in structured additive regression models for survival data.
- Koop, G. and D. Korobilis (2012). Forecasting inflation using dynamic model averaging. *International Economic Review* 53(3), 867–886.
- Koop, G., S. McIntyre, J. Mitchell, and A. Poon (2020). Regional output growth in the united kingdom: More timely and higher frequency estimates from 1970. *Journal of Applied Econometrics* 35(2), 176–197.
- 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.
- Korobilis, D. (2017). Quantile regression forecasts of inflation under model uncertainty. *International Journal of Forecasting* 33(1), 11–20.
- Kowal, D. R. (2021). Fast, optimal, and targeted predictions using parametrized decision analysis. *Journal of the American Statistical Association*, 1–28.
- Kowal, D. R., D. S. Matteson, and D. Ruppert (2019). Dynamic shrinkage processes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 81(4), 781–804.
- Kozumi, H. and G. Kobayashi (2011). Gibbs sampling methods for bayesian quantile regression. *Journal of statistical computation and simulation* 81(11), 1565–1578.
- Lee, E. R., H. Noh, and B. U. Park (2014). Model selection via bayesian information criterion for quantile regression models. *Journal of the American Statistical Association* 109(505), 216–229.
- Lenza, M. and G. E. Primiceri (2020). How to estimate a var after march 2020. Technical report, National Bureau of Economic Research.
- Lenza, M. and G. E. Primiceri (2022). How to estimate a vector autoregression after march 2020. *Journal of Applied Econometrics*.
- Li, Q., R. Xi, N. Lin, et al. (2010). Bayesian regularized quantile regression. *Bayesian Analysis* 5(3), 533–556.

- Lindley, D. V. (1968). The choice of variables in multiple regression. *Journal of the Royal Statistical Society: Series B (Methodological)* 30(1), 31–53.
- Litterman, R. B. (1980). *Bayesian procedure for forecasting with vector autoregressions*. Massachusetts Institute of Technology.
- Makalic, E. and D. F. Schmidt (2015). A simple sampler for the horseshoe estimator. *IEEE Signal Processing Letters* 23(1), 179–182.
- Malsiner-Walli, G. and H. Wagner (2018). Comparing spike and slab priors for bayesian variable selection. *arXiv preprint arXiv:1812.07259*.
- 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.
- Mariano, R. S. and Y. Murasawa (2003). A new coincident index of business cycles based on monthly and quarterly series. *Journal of applied Econometrics* 18(4), 427–443.
- Mazzi, G. L. and J. Mitchell (2019). Nowcasting euro area gdp growth using quantile regression.
- Mazzi, G. L., J. Mitchell, and G. Montana (2014). Density nowcasts and model combination: Nowcasting euro-area gdp growth over the 2008–09 recession. *Oxford Bulletin of Economics and Statistics* 76(2), 233–256.
- McAlinn, K., V. Rockova, and E. Saha (2018). Dynamic sparse factor analysis. *arXiv preprint arXiv:1812.04187*.
- McAlinn, K. and M. West (2019). Dynamic bayesian predictive synthesis in time series forecasting. *Journal of econometrics* 210(1), 155–169.
- McCausland, W. J., S. Miller, and D. Pelletier (2011). Simulation smoothing for state–space models: A computational efficiency analysis. *Computational Statistics & Data Analysis* 55(1), 199–212.
- McConnell, M. M. and G. Perez-Quiros (2000). Output fluctuations in the united states: What has changed since the early 1980’s? *American Economic Review* 90(5), 1464–1476.
- McCracken, M. and S. Ng (2020). Fred-qd: A quarterly database for macroeconomic research. Technical report, National Bureau of Economic Research.

- McCracken, M. W., J. T. McGillicuddy, and M. T. Owyang (2021). Binary conditional forecasts. *Journal of Business & Economic Statistics*, 1–13.
- McCracken, M. W. and S. Ng (2016). Fred-md: A monthly database for macroeconomic research. *Journal of Business & Economic Statistics* 34(4), 574–589.
- McLaren, N. and R. Shanbhogue (2011). Using internet search data as economic indicators. *Bank of England Quarterly Bulletin* (2011), Q2.
- Mitchell, J., A. Poon, and G. L. Mazzi (2022). Nowcasting euro area gdp growth using bayesian quantile regression. In *Essays in Honor of M. Hashem Pesaran: Prediction and Macro Modeling*. Emerald Publishing Limited.
- Mitchell, T. J. and J. J. Beauchamp (1988). Bayesian variable selection in linear regression. *Journal of the american statistical association* 83(404), 1023–1032.
- Mogliani, M. and A. Simoni (2021). Bayesian midas penalized regressions: estimation, selection, and prediction. *Journal of Econometrics* 222(1), 833–860.
- Moran, G. E., V. Ročková, E. I. George, et al. (2018). Variance prior forms for high-dimensional bayesian variable selection. *Bayesian Analysis*, 1091–1119.
- Ng, S. (2013). Variable selection in predictive regressions. In *Handbook of economic forecasting*, Volume 2, pp. 752–789. Elsevier.
- Ng, S. (2021). Modeling macroeconomic variations after covid-19. Technical report, National Bureau of Economic Research.
- 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.
- O’Hara, R. B. and M. J. Sillanpää (2009). A review of bayesian variable selection methods: what, how and which. *Bayesian analysis* 4(1), 85–117.
- Park, T. and G. Casella (2008). The bayesian lasso. *Journal of the American Statistical Association* 103(482), 681–686.
- Pavone, F., J. Piironen, P.-C. Bürkner, and A. Vehtari (2020). Using reference models in variable selection. *arXiv preprint arXiv:2004.13118*.
- Piironen, J., M. Paasiniemi, A. Vehtari, et al. (2020). Projective inference in high-dimensional problems: Prediction and feature selection. *Electronic Journal of Statistics* 14(1), 2155–2197.

- Piironen, J. and A. Vehtari (2017). Comparison of bayesian predictive methods for model selection. *Statistics and Computing* 27(3), 711–735.
- Piironen, J., A. Vehtari, et al. (2017). Sparsity information and regularization in the horseshoe and other shrinkage priors. *Electronic Journal of Statistics* 11(2), 5018–5051.
- Polson, N. G. and J. G. Scott (2010). Shrink globally, act locally: Sparse bayesian regularization and prediction. *Bayesian statistics* 9, 501–538.
- 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.
- Prasad, M. A., S. Elekdag, M. P. Jeasakul, R. Lafarguette, M. A. Alter, A. X. Feng, and C. Wang (2019). *Growth at risk: Concept and application in imf country surveillance*. International Monetary Fund.
- Prüser, J. (2019). Forecasting with many predictors using bayesian additive regression trees. *Journal of Forecasting* 38(7), 621–631.
- Raftery, A. E., D. Madigan, and J. A. Hoeting (1997). Bayesian model averaging for linear regression models. *Journal of the American Statistical Association* 92(437), 179–191.
- Ray, P. and A. Bhattacharya (2018). Signal adaptive variable selector for the horseshoe prior. *arXiv preprint arXiv:1810.09004*.
- Roberts, S., M. Osborne, M. Ebden, S. Reece, N. Gibson, and S. Aigrain (2013). Gaussian processes for time-series modelling. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 371(1984), 20110550.
- Rockova, V. and K. McAlinn (2021). Dynamic variable selection with spike-and-slab process priors. *Bayesian Analysis* 16(1), 233–269.
- Romer, D. (2018). *Advanced macroeconomics*. Mcgraw-hill.
- Rossi, B. and T. Sekhposyan (2019). Alternative tests for correct specification of conditional predictive densities. *Journal of Econometrics* 208(2), 638–657.
- Schorfheide, F. and D. Song (2015). Real-time forecasting with a mixed-frequency var. *Journal of Business & Economic Statistics* 33(3), 366–380.
- Schorfheide, F. and D. Song (2021). Real-time forecasting with a (standard) mixed-frequency var during a pandemic. Technical report, National Bureau of Economic Research.

- 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.
- Scruton, J., M. O’Donnell, and S. Dey-Chowdhury (2018). Introducing a new publication model for gdp. *Office for National Statistics*, 1–14.
- Shapiro, A. H., M. Sudhof, and D. J. Wilson (2020). Measuring news sentiment. *Journal of Econometrics*.
- Shin, M. and J. S. Liu (2021). Neuronized priors for bayesian sparse linear regression. *Journal of the American Statistical Association*, 1–16.
- Simon, N. and R. Tibshirani (2012). Standardization and the group lasso penalty. *Statistica Sinica* 22(3), 983.
- 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.
- Smith, R. G. and D. E. Giles (1976). *The Almon estimator: Methodology and users’ guide*. Reserve Bank of New Zealand.
- Sriram, K., R. Ramamoorthi, and P. Ghosh (2013). Posterior consistency of bayesian quantile regression based on the misspecified asymmetric laplace density. *Bayesian analysis* 8(2), 479–504.
- Stock, J. H. and M. W. Watson (1998). A comparison of linear and nonlinear univariate models for forecasting macroeconomic time series.
- Stock, J. H. and M. W. Watson (2002). Forecasting using principal components from a large number of predictors. *Journal of the American statistical association* 97(460), 1167–1179.
- Stock, J. H. and M. W. Watson (2004). Combination forecasts of output growth in a seven-country data set. *Journal of forecasting* 23(6), 405–430.
- Stock, J. H. and M. W. Watson (2007). Why has us inflation become harder to forecast? *Journal of Money, Credit and banking* 39, 3–33.
- Stock, J. H. and M. W. Watson (2012a). Disentangling the channels of the 2007-2009 recession. Technical report, National Bureau of Economic Research.
- Stock, J. H. and M. W. Watson (2012b). Generalized shrinkage methods for forecasting using many predictors. *Journal of Business & Economic Statistics* 30(4), 481–493.

- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)* 58(1), 267–288.
- Van Der Pas, S. L., B. J. Kleijn, A. W. Van Der Vaart, et al. (2014). The horseshoe estimator: Posterior concentration around nearly black vectors. *Electronic Journal of Statistics* 8(2), 2585–2618.
- Varian, H. R. (2014). Big data: New tricks for econometrics. *Journal of Economic Perspectives* 28(2), 3–28.
- Vehtari, A., A. Gelman, and J. Gabry (2017). Practical bayesian model evaluation using leave-one-out cross-validation and waic. *Statistics and computing* 27(5), 1413–1432.
- Vehtari, A. and J. Ojanen (2012). A survey of bayesian predictive methods for model assessment, selection and comparison. *Statistics Surveys* 6, 142–228.
- 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.
- Wang, H. and C. Leng (2008). A note on adaptive group lasso. *Computational statistics & data analysis* 52(12), 5277–5286.
- Wang, J. (2012). Bayesian quantile regression for parametric nonlinear mixed effects models. *Statistical Methods & Applications* 21(3), 279–295.
- Wang, L. (2017). Nonconvex penalized quantile regression: A review of methods, theory and algorithms. In *Handbook of Quantile Regression*, pp. 273–292. Chapman and Hall/CRC.
- Watson, M. W. (1986). Univariate detrending methods with stochastic trends. *Journal of monetary economics* 18(1), 49–75.
- Welch, I. and A. Goyal (2008). A comprehensive look at the empirical performance of equity premium prediction. *The Review of Financial Studies* 21(4), 1455–1508.
- Williams, C. K. and C. E. Rasmussen (2006). *Gaussian processes for machine learning*, Volume 2. MIT press Cambridge, MA.
- 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.

- Woody, S., C. M. Carvalho, and J. S. Murray (2020). Model interpretation through lower-dimensional posterior summarization. *Journal of Computational and Graphical Statistics*, 1–9.
- Woody, S., C. M. Carvalho, and J. S. Murray (2021). Model interpretation through lower-dimensional posterior summarization. *Journal of Computational and Graphical Statistics* 30(1), 144–161.
- 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.
- Xu, X. and M. Ghosh (2015). Bayesian variable selection and estimation for group lasso. *Bayesian Analysis* 10(4), 909–936.
- Yang, Y., H. J. Wang, and X. He (2016). Posterior inference in bayesian quantile regression with asymmetric laplace likelihood. *International Statistical Review* 84(3), 327–344.
- Yu, K. and R. A. Moyeed (2001). Bayesian quantile regression. *Statistics & Probability Letters* 54(4), 437–447.
- Zhang, Y. D., B. P. Naughton, H. D. Bondell, and B. J. Reich (2020). Bayesian regression using a prior on the model fit: The r2-d2 shrinkage prior. *Journal of the American Statistical Association* (just-accepted), 1–37.
- Zhang, Y. D., B. P. Naughton, H. D. Bondell, and B. J. Reich (2022). Bayesian regression using a prior on the model fit: The r2-d2 shrinkage prior. *Journal of the American Statistical Association* 117(538), 862–874.
- Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American statistical association* 101(476), 1418–1429.