

On the use of multiple models in macro-economic  
forecasting and decision-making

PhD Thesis

Tony Chernis Department of Economics  
University of Strathclyde, Glasgow

August 9, 2024

This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination which has led to the award of a degree.

The copyright of this thesis belongs to the author under the terms of the United Kingdom Copyright Acts as qualified by University of Strathclyde Regulation 3.50. Due acknowledgement must always be made of the use of any material contained in, or derived from, this thesis.

# Acknowledgements

First, I would like to thank my supervisors Gary Koop and Stuart McIntyre. Gary for patiently responding to my constant barrage of emails and questions while also giving me the opportunity to work with him. Stuart has been immensely helpful with many research discussions, advice, and has expertly guided me through the PhD process.

I would also like to thank Mike West. I was lucky enough to co-author with Mike which taught me a lot about doing research, statistical thinking, and Bayesian methods. Additionally, I would like to thank my co-authors Niko Hauzenberger, Florian Huber, James Mitchell, and Emily Tallman. These collaborations were a formative experience that has contributed greatly to my development as an econometrician.

I would also like to thank Rodrigo Sekkel. Rodrigo has for the last ten years been a terrific mentor, advisor, and friend. I am not sure where I would be without his help.

Many thanks are also in order for Paul Beaudry. Paul has graciously spent hours chatting with me about economics and research. I also owe him thanks for encouraging me to pursue this PhD. It was his advice which finally convinced me to take the plunge.

I would also like to thank Luis Uzeda for being so helpful and always taking the time to comment on my papers. He consistently provided useful research direction. I am also grateful to Saeed Zaman for being very positive and encouraging me to pursue the PhD at Strathclyde.

I gratefully acknowledge the funding and support for my PhD from the Bank of Canada. Specifically, I would like to thank Calista Cheung, Joshua Slive, and Brigitte Desroches for giving me the time and support necessary to complete my PhD. Without their support this would not have been possible.

I would like to thank my Mom, Deborah Wilkins, for always believing in me and

encouraging me to follow my dreams. And, finally, I would like to thank my wife, Tricia Riley, for the constant unwavering support. She was always there to listen to me during the program and her proof-reading has made my papers readable.

# Abstract

This thesis studies prediction and decision-making with multiple models in a Bayesian context. It is common practice to use multiple models in a forecasting and decision-making environment. This is because the true model is unknown, if it even exists, and even the ‘best’ model can be hard to identify. Consequently, there is often large amounts of uncertainty around the choice of model and decision-makers resort to multiple models. Using multiple models in practice is an open area of research where this thesis will contribute over the course of three essays. The three problems I seek to address are how to combine large numbers of forecasts; how to explain why combination techniques place higher, or lower, weight on certain models; and finally, how to use multiple models in a monetary policy decision-making context.

The first essay (Chapter 2) investigates model combination with large numbers of models and predictions. To this end I use Bayesian Predictive Synthesis (BPS) which is a flexible method of combining density predictions. The flexibility comes from the ability to choose an arbitrary synthesis function to combine predictions. Through careful choice of this synthesis function I show how to combine large numbers of predictions—which is a common occurrence in macroeconomics. Specifically, I consider shrinkage priors and factor modeling techniques which are common choices for high-dimensional problems in macroeconomics. Additionally, these techniques provide an interesting contrast between the sparse weights implied by shrinkage priors and dense weights of factor modeling techniques. I find that the sparse weights of shrinkage priors perform well across exercises.

The second essay (Chapter 3) addresses a common issue which is that it can be difficult to understand the reason why models are chosen in a combination. This is

of particular importance in decision-making contexts. As in Chapter 2 we develop a synthesis function to address this problem. Typically, synthesis functions are specified parametrically as a dynamic linear regression. Instead, we develop a nonparametric treatment of the synthesis function using regression trees. We are able to explain the combination weights since we introduce observable variables in the regressions trees' splitting rule. We can then examine the tree splits to see which variables are most important for explaining the weights. We show the advantages of our tree-based approach in two macroeconomic forecasting applications. The first uses density forecasts for GDP growth from the euro area's Survey of Professional Forecasters. The second combines density forecasts of US inflation produced by many regression models involving different predictors. Both applications demonstrate the benefits – in terms of improved forecast accuracy and interpretability – of modeling the synthesis function nonparametrically.

The third essay (Chapter 4) goes beyond the previous chapters going from prediction to decision-making. We show to make optimal monetary policy decisions with multiple models. We use Bayesian predictive decision synthesis (BPDS) as a formal Bayesian decision theory-based approach to monetary policy decision-making. BPDS draws on recent developments in model combination and statistical decision theory that make it possible to combine models in a manner that incorporates decision goals, expectations, and outcomes. We develop a BPDS procedure for a case study of monetary policy decision-making with an inflation-targeting central bank. Our procedure searches for an optimal monetary policy decision through maximizing the decision-maker's utility function and weighting models conditionally on that decision. The model weights are determined by their empirical fit, past and expected decision-making performance, and the model-based plausibility of the policy decision. We find that BPDS produces quite different decisions and weights from standard approaches, such as Bayesian Model Averaging, that only consider forecasting performance.

# Contents

<b>List of Figures</b>	<b>ix</b>
<b>List of Tables</b>	<b>xii</b>
<b>1 Introduction</b>	<b>2</b>
1.1 Motivation and Contribution . . . . .	2
<b>2 Combining Large Numbers of Density Predictions with Bayesian Predictive Synthesis</b>	<b>6</b>
2.1 Introduction . . . . .	6
2.2 Econometric Framework . . . . .	10
2.2.1 Bayesian Predictive Synthesis . . . . .	10
2.2.2 Global-local Shrinkage Priors . . . . .	12
2.2.3 Factor-model-based Combination . . . . .	13
2.3 Forecasting Environment . . . . .	14
2.3.1 Details on the Model-based Nowcasting Exercise . . . . .	15
2.3.2 Details on the Survey of Professional Forecasters . . . . .	17
2.4 Results . . . . .	18
2.4.1 Predictive Accuracy: Global-local Shrinkage Priors and Factor Model Combinations . . . . .	19
2.4.2 Predictive Accuracy: Time-varying and Constant Parameter Combinations . . . . .	25
2.4.3 Examining the Combination Weights . . . . .	26
2.5 Conclusion . . . . .	28

<b>3</b>	<b>Predictive Density Combination Using a Tree-Based Synthesis Function</b>	<b>30</b>
3.1	Introduction . . . . .	30
3.2	Bayesian Predictive Synthesis with Regression Trees . . . . .	34
3.2.1	Bayesian Predictive Synthesis . . . . .	34
3.2.2	Bayesian Predictive Synthesis with Regression Trees (BPS-RT) . . . . .	38
3.2.3	Illustrating BPS-RT . . . . .	42
3.3	Two Macroeconomic Forecasting Applications . . . . .	45
3.3.1	BPS-RT Specifications . . . . .	46
3.3.2	Forecasting EA Output Growth Using the Survey of Professional Forecasters . . . . .	49
3.3.3	Forecasting US inflation Using a Set of Indicators from FRED-QD . . . . .	51
3.3.4	Empirical Results . . . . .	52
3.4	Conclusion . . . . .	66
<b>4</b>	<b>Decision Synthesis in Monetary Policy</b>	<b>68</b>
4.1	Introduction . . . . .	68
4.2	BPDS Framework . . . . .	71
4.2.1	Mixture BPDS and Decision Setting . . . . .	71
4.2.2	Decision-dependent Scores for Calibration Functions . . . . .	74
4.2.3	BPDS Summary . . . . .	75
4.3	BPDS for Optimal Monetary Policy Decisions . . . . .	76
4.3.1	Models, Forecasts, and Model-specific Decisions . . . . .	76
4.3.2	BPDS Model Specification . . . . .	78
4.3.3	BPDS Implementation and Optimal Decisions . . . . .	83
4.4	Case Study . . . . .	86
4.4.1	Overview . . . . .	86
4.4.2	Optimal Decisions . . . . .	86
4.4.3	Trajectories of BPDS and BMA Predictive Densities . . . . .	88
4.4.4	Model Probabilities . . . . .	89
4.4.5	Additional Insights from BPDS Results . . . . .	94

## Contents

4.5	Summary Comments . . . . .	96
<b>5</b>	<b>Conclusion</b>	<b>99</b>
5.1	Summary Comments . . . . .	99
5.2	Further Research . . . . .	103
<b>A</b>	<b>Chapter 2 Appendix</b>	<b>104</b>
A.1	Tables . . . . .	104
A.2	Technical Appendix . . . . .	105
A.2.1	MCMC Algorithm . . . . .	105
A.2.2	Global-local Shrinkage Combinations . . . . .	105
A.2.3	Factor Model Combinations . . . . .	110
A.2.4	Sampling the Agent States . . . . .	111
A.3	Calibration Appendix . . . . .	113
A.4	MCMC Convergence Appendix . . . . .	115
<b>B</b>	<b>Chapter 3 Appendix</b>	<b>117</b>
B.1	Technical Appendix: Bayesian Inference . . . . .	118
B.1.1	Additional Details about Our Modeling Choices and Priors . . . . .	118
B.1.2	Posterior Simulation . . . . .	121
B.1.3	ADL Model Estimation . . . . .	125
B.2	Empirical Appendix: Additional Results . . . . .	126
B.2.1	Tail forecast accuracy . . . . .	126
B.2.2	Combination Weights . . . . .	128
B.2.3	Cumulative CRPS . . . . .	132
B.2.4	Giacomini and Rossi (2010) Fluctuation Test Statistic . . . . .	134
B.2.5	Probability Integral Transforms (PITs) . . . . .	135
B.2.6	Measuring the Degree of Shrinkage . . . . .	136
B.2.7	Combined Predictive Densities and Their Skewness . . . . .	137
<b>C</b>	<b>Chapter 4 Appendix</b>	<b>139</b>
C.1	Evaluation of Tilting Vectors . . . . .	140

Contents

C.2	Conditional Forecasts . . . . .	141
C.2.1	Observational Constraints . . . . .	141
C.2.2	Constraints on Structural Shocks . . . . .	143
	<b>Bibliography</b>	<b>144</b>

# List of Figures

2.1	Model List . . . . .	15
2.2	Overview of Forecast Cycle . . . . .	16
2.3	Comparison of Shrinkage Priors and Factor Models . . . . .	20
2.4	Comparison of Time-varying and Constant Parameter Specification . . . . .	20
2.5	Cumulative CRPS Difference . . . . .	21
2.6	Nowcasting Application Predictive Densities . . . . .	22
2.7	SPF Application Predictive Densities . . . . .	22
2.8	Nowcasting Application Difference Between Triplegamma and Factor Model-2 in Predictive Densities with GDP Realizations . . . . .	23
2.9	SPF Application: Difference Between Triple Gamma and Factor Model-2 in Predictive Densities with GDP Realizations . . . . .	24
2.10	Time-varying Predictive of Mean of BPS Intercepts for SPF Tall Dataset . . . . .	26
2.11	Sequentially Estimated Mean Combination Weights . . . . .	27
3.1	Illustration of BPS-RT. . . . .	44
3.2	Relative forecast accuracy: EA GDP growth. . . . .	55
3.3	Relative forecast accuracy: US inflation. . . . .	56
3.4	Difference in probabilities between BPS-RT and BPS-RW . . . . .	59
3.5	Measuring model incompleteness: US inflation . . . . .	61
3.6	Number of tree splits for BPS-RT ( $S = 1$ ) and relative importance for each weight modifier for US inflation: One-quarter-ahead forecasts . . . . .	62
3.7	<b>One-year-ahead horizon:</b> Number of (total) tree splits for our single-tree models and relative importance for each effect modifier. . . . .	65

## List of Figures

4.1	Recursively calculated policy decisions . . . . .	88
4.2	BPDS forecast densities of inflation. The frames represent 1–8 quarter ahead forecasts, reading along the rows from top-left to bottom-right. The colours represent probabilities, with the blue shading showing lower probability, and red showing higher probability. . . . .	90
4.3	BMA forecast densities of inflation. The frames represent 1–8 quarter ahead forecasts, reading along the rows from top-left to bottom-right. The colours represent probabilities, with the blue shading showing lower probability, and red showing higher probability. . . . .	91
4.4	Difference between BPDS and BMA forecast densities of inflation. Red-shaded regions have higher probability under BPDS than under BMA, with blue shading indicating the reverse, and white shading showing equal probability. . . . .	92
4.5	Time trajectories of model probabilities. (a) Prior BPDS probabilities $\pi_{tj}$ based on discounted AVS with a fixed baseline $\pi_{t0} = 0.1$ ; (b) BPDS decision-dependent initial probabilities $\pi_{tj}(\mathbf{x}_t)$ ; (c) Implied BPDS weights $\tilde{\pi}_{tj}(\mathbf{x}_t)$ ; (d) BMA probabilities. . . . .	93
4.6	Trajectories of the eight elements of the evaluated BPDS tilting vector $\boldsymbol{\tau}_t(\mathbf{x}_t)$ at the optimized $\mathbf{x}_t$ at each quarter. . . . .	94
4.7	Trajectories of the effective sample size (ESS) metrics for individual models and for the BPDS initial mixture. . . . .	94
4.8	Trajectories of expected utilities comparing BPDS with BMA. . . . .	96
A.1	Knüpple Test for Probabilistic Calibration: Nowcasting Application . . .	114
A.2	Knüpple Test for Probabilistic Calibration: Survey Forecast Application	115
A.3	Box Plots of Potential Scale Reduction Factors . . . . .	116
B.1	Relative tail forecast accuracy: EA GDP growth. . . . .	126
B.2	Relative tail forecast accuracy: US inflation. . . . .	127
B.3	Combination weights over the evaluation sample: EA GDP growth . . .	128

## List of Figures

B.4	Combination weights over the evaluation sample: One-quarter-ahead US inflation . . . . .	129
B.5	Combination weights over the evaluation sample: One-year-ahead US inflation ( $h = 4$ ) . . . . .	130
B.6	Sum of combination weights over the evaluation sample . . . . .	131
B.7	Forecast performance of single-tree specifications with stochastic volatility: EA GDP growth . . . . .	132
B.8	Forecast performance of single-tree specifications with homoskedastic error variances: US inflation . . . . .	133
B.9	Evolution of the Giacomini and Rossi (2010) fluctuation test statistic . .	134
B.10	Evaluating model calibration using probability integral transforms (PITs)	135
B.11	Overall degree of shrinkage toward the prior mean for US inflation. . .	136
B.12	BPS-RT and BPS-RW predictive densities . . . . .	137
B.13	Evolution of a quantile-based skewness measure for the BPS-RT predictive densities . . . . .	138

# List of Tables

A.1	Nowcasting Application: Overview of Forecasting Performance . . . . .	104
A.2	SPF Application: Overview of Forecasting Performance . . . . .	105
A.3	The rows show results for wide and tall datasets. The columns correspond to the Dynamic Linear Model benchmark (DLM), constant and time-varying specification of the Lasso, double gamma prior (DG), triple gamma prior (TG), Horseshoe prior (HS), and factor model synthesis functions with 1 to 5 factors. . . . .	105
B.1	List of variables used for the autoregressive distributed lag (ADL) specifications. . . . .	125

## List of Tables

# Chapter 1

## Introduction

*‘All models are wrong but some are useful’*

George Box

### 1.1 Motivation and Contribution

The preceding aphorism from George Box underlies much of the thinking in this thesis and much of macroeconomic modelling. It is a very simply, yet insightful, observation that the models are an abstraction from reality and are thus imperfect. However, it is this abstraction that allows a model to draw out insight. Compounding this problem is that for policy-makers and practitioners it may not even be clear which model is best suited for their task. Consequently, it is common practice to use multiple models in a forecasting and decision-making environment to hedge against model uncertainty.

The case for combining forecasts in economics goes back at least as far as Bates and Granger (1969). While Timmermann (2006a) provides a convincing motivation based on hedging against model uncertainty. The basis of the arguments for model combination comes from that observation that by simply choosing the best performing model one might be throwing out useful information contained in the other candidate models. This becomes even more complicated once you consider that choosing the best model is not always easy. The argument then follows that a practitioner who is concerned with making the best forecast or decision should combine the information in all the models. Furthermore, to hedge against choosing a poorly performing model, a

practitioner should average across candidate models. Judged by their uptake by Central Banks<sup>1</sup> forecast combination techniques have been proven a successful strategy. While Aastveit *et al.* (2019) provides a modern overview of density combination in economics which recounts many successes of density combinations. Despite the proliferation of research, density combinations are still an activity area of research and this thesis investigates several issues present when using multiple models within a Bayesian context. Broadly speaking this thesis investigates the choice of functional forms to combine forecasts, how to explain the weighting of models, and how to make decisions with these models.

The first essay (Chapter 2) investigates model combination with large numbers of models and predictions. This is a common scenario in macro-economic contexts. For example, surveys of forecasters can include dozens of respondents while some nowcasting systems have close to a hundred models. With large numbers of models it can be challenging to find an appropriate weighting for each model due to the relatively short-sample size in macroeconomic data. To this end I use Bayesian Predictive Synthesis (BPS) which is a flexible method of combining density predictions. The flexibility comes from the ability to choose an arbitrary synthesis function to combine predictions. Through careful choice of this synthesis function I show how to combine large numbers of predictions. Specifically, I consider shrinkage priors and factor modeling techniques which are common choices for high-dimensional problems in macroeconomics. Additionally, these techniques provide an interesting contrast between the sparse weights implied by shrinkage priors and dense weights of factor modeling techniques. I find that the sparse weights of shrinkage priors perform well across exercises and simple constant weight specifications perform best.

The second essay (Chapter 3) addresses a common issue which is that it can be difficult to understand the reason why models are chosen in a combination. This is of particular importance in decision-making contexts. It can be tempting to hand-wave this away by appealing to forecast accuracy metrics, but this can be unsatisfying for

---

<sup>1</sup>For example, the Bank of Canada (Chernis and Sekkel, 2018; Chernis and Webley, 2022), the Norges Bank (Bjørnland *et al.*, 2012; Aastveit *et al.*, 2011), and the Bank of England (Kapetanios *et al.*, 2008)

## Chapter 1. Introduction

a policy-maker when they are forced to defend their forecast. As in Chapter 2 we develop a synthesis function to address this problem. Typically, the synthesis function is specified parametrically as a dynamic linear regression. Instead, we develop a nonparametric treatment of the synthesis function using regression trees. We are able explain the combination weights since we introduce observable variables in the regressions trees splitting rule. We can then examine the tree splits to see which variables are most important for explaining the weights. We show the advantages of our tree-based approach in two macroeconomic forecasting applications. The first uses density forecasts for GDP growth from the euro area’s Survey of Professional Forecasters. The second combines density forecasts of US inflation produced by many regression models involving different predictors. Both applications demonstrate the benefits – in terms of improved forecast accuracy and interpretability – of modeling the synthesis function nonparametrically.

The third essay (Chapter 4) addresses an often neglected issues which is that forecasts are often used to make a decision. I show how to make optimal monetary policy decisions with multiple models while weighting models with the goal of achieving an inflation target in mind. We use Bayesian predictive decision synthesis (BPDS) as a formal Bayesian decision-theory based approach to monetary policy decision-making. BPDS draws on recent developments in model combination and statistical decision theory that make it possible to combine models in a manner that incorporates decision goals, expectations, and outcomes. We develop a BPDS procedure for a case study of monetary policy decision-making with an inflation-targeting central bank. Our procedure searches for an optimal monetary policy decision through maximizing a decision-maker’s utility function then weighting models conditional on that decision. The models considered by the decision-maker are determined by their empirical fit, past and expected decision making performance, and the model-based plausibility of the policy decision. We find that BPDS produces quite different decisions and weights from standard approaches, such as Bayesian Model Averaging, that only consider forecasting performance.

As of July 2024, Chapter 2 has been published as Chernis (2024). A working paper version of Chapter 3 has been published as Chernis *et al.* (2023) and is submitted for possible publication. Chapter 3 is joint work with Gary Koop, Niko Hauzenberger,

## Chapter 1. Introduction

Florian Huber, and James Mitchell. Similarly, a working paper version of Chapter 4, which is co-authored with Gary Koop, Emily Tallman, and Mike West, has been published as Chernis *et al.* (2024) and is submitted for possible publication.

## Chapter 2

# Combining Large Numbers of Density Predictions with Bayesian Predictive Synthesis

### 2.1 Introduction

This paper develops and studies several techniques to combine large numbers of predictive densities. This is a common problem since decision-makers often consult a wide variety of models and experts to form the basis of their decision-making (Coletti and Murchison, 2002). They consult multiple models because of the recognition that individual models (or experts) often provide a partial understanding of the economy due to different underlying datasets or modeling assumptions, creating significant uncertainty around their predictions. It is useful for decision-makers to understand the uncertainty around a prediction. Therefore, practitioners create predictive densities and combine them to characterize uncertainty from individual predictions and model choice (Chernis and Webley, 2022). These density combinations show not just the uncertainty around a prediction, but the balance of risks, or the severity of tail risks. However, combining density predictions often involves large numbers of densities, such as in nowcasting platforms or expert surveys, which can be a difficult task and requires specialized techniques.

## Chapter 2. Combining Large Numbers of Density Predictions with Bayesian Predictive Synthesis

I address the issue of combining large numbers of density forecasts using two approaches that are commonly used to deal with large datasets in economics. Specifically, I compare global-local shrinkage priors and factor models when combining predictive densities within the framework of Bayesian Predictive Synthesis (BPS). Global-local shrinkage priors and factor models naturally lend themselves to large dimensional problems. BPS is an approach for combining densities that allows the user to choose the functional form by which the predictions will be combined. In particular, I use the triple gamma prior (Cadonna *et al.*, 2020) as a baseline global-local shrinkage prior since it nests many commonly used priors, such as the horseshoe (Carvalho *et al.*, 2010) and the Bayesian Lasso (Belmonte *et al.*, 2014). Because of this feature, I also compare the performance of various hierarchical priors. Additionally, I develop a Bayesian Factor Model (Lopes, 2014) to combine forecasts. To the best of my knowledge, this is a novel method of combining density predictions. The closest approach I am aware of is Casarin *et al.* (2019), who model the weights as correlated using a factor structure. In contrast, in this paper the forecasts are modeled as correlated and have a factor structure.

I find that global-local shrinkage priors generally outperform factor models as measured by the Continuous Rank Probability Score (CRPS) of Gneiting and Raftery (2007). Since shrinkage priors induce sparsity, this finding suggests that focusing on a smaller set of accurate experts is preferable to following the herd. Another important finding relates to the specification of the synthesis functions: I find that constant parameter models are a more reliable choice. The extra flexibility from allowing time-varying weights can cause the accuracy of the forecasts to deteriorate significantly. In addition, in some cases, time-varying parameter specifications can reduce to a time-varying mean model that overfits the model, resulting in poor out-of-sample performance.

This kind of analysis is only possible in a BPS framework. BPS frames the issue of combining predictions as a decision theory problem—a decision-maker rationally synthesizes some set of information to inform their choice of action. The theoretical underpinnings are provided by West and Crosse (1992) and West (1992), who show how a decision-maker would combine a set of forecast distributions (or partial summaries) in a fully Bayesian manner. Recently, this has been codified by McAlinn and West (2019),

## Chapter 2. Combining Large Numbers of Density Predictions with Bayesian Predictive Synthesis

who introduce Bayesian Predictive Synthesis for time series. Apart from the strong theoretical motivation for using BPS, it is very flexible. A researcher can specify the functional form, called the synthesis function, of the density combination with very few restrictions. This makes it very easy to compare and experiment with different ways of combining forecasts.

So far, comparisons of synthesis functions have not been addressed in a BPS framework. Most applications of BPS use a dynamic linear model as a synthesis function (Prado and West, 2010, Sect. 4.5). Instead, BPS is extended to a multivariate forecast setting in McAlinn *et al.* (2020). Takanashi and McAlinn (2021) establish additional theoretical properties such as the BPS combined predictions being minimax. McAlinn (2021) uses BPS in a mixed-frequency nowcasting exercise, and Aastveit *et al.* (2023) use it to forecast oil prices.

Comparing global-local shrinkage priors and factor-model-based synthesis functions is interesting for several reasons. These approaches naturally allow for combining forecasts with a large number of experts. This is significant since many applications feature large numbers of experts, such as nowcasting with ensembles and survey forecasts. This can be challenging due to the requirement of estimating a large number of parameters with small datasets. From a frequentist perspective, the approach is to employ regularization while estimating an optimal combination (Conflitti *et al.*, 2015; Diebold *et al.*, 2023). Bayesian approaches can also face difficulties in large dimensions. For example, Bayesian Model Averaging requires calculation of the marginal likelihood for each model, which is computationally expensive. Researchers have addressed this issue using approximations (Jore *et al.*, 2010) or reducing the number of marginal likelihoods to be calculated (Onorante and Raftery, 2016). Another solution is to estimate clusters of weights instead of weights for each individual model, such as in Billio *et al.* (2013) and Casarin *et al.* (2019).

Additionally, global-local shrinkage priors and factor models have very different properties. Shrinkage priors tend to place weight on a smaller subset of experts (sparsity), while factor models look for co-movement and the weights are more egalitarian (or dense). To the best of my knowledge, this contrast has not been examined in the

## Chapter 2. Combining Large Numbers of Density Predictions with Bayesian Predictive Synthesis

density forecast combination literature. This is in contrast to studies that examine whether a dense representation of the data is appropriate (Giannone *et al.*, 2021; Cross *et al.*, 2020) or an artifact of prior choice (Fava and Lopes, 2021). In the context of density combinations, this is equivalent to asking, “should a decision-maker pick winners or follow the herd?” when provided with views on the economy.

This paper is part of a long history of research on forecast combinations in macroeconomics, econometrics, and statistics. Over the past twenty years, a lot of progress has been made in the study of density combinations in economics.<sup>1</sup> Several authors show that combining densities can make predictions more robust and improve their accuracy (Jore *et al.*, 2010; Del Negro *et al.*, 2016), while others specify optimal combination strategies from both frequentist (Conflitti *et al.*, 2015) and Bayesian perspectives (Geweke and Amisano, 2011). More recent academic work focuses on modeling the dependence and correlation across forecasts, and time variation in weights.<sup>2</sup> Furthermore, Knotek and Zaman (2023) and Chernis and Webley (2022) show how density combinations can have non-Gaussian and time-varying features, which improves the predictions and are useful for characterizing uncertainty. Similar to point forecast combinations, density combinations have also proven useful in central banks (Bjørnland *et al.*, 2012; Aastveit *et al.*, 2011). For a thorough review of the evolution of density predictions in economics and its advantages, see Aastveit *et al.* (2019).

The remainder of the paper proceeds as follows. Section 2.2 describes Bayesian Predictive Synthesis along with an outline of the Markov Chain Monte Carlo (MCMC) approach. This is followed by a description of the forecast combination techniques—the synthesis functions. In addition, I provide a brief overview of global-local shrinkage priors and Bayesian factor models. Section 2.3 details the prediction exercises, while Section 2.4 discusses the results. Section 2.5 concludes.

---

<sup>1</sup>For example, Wallis (2005); Hall and Mitchell (2007); Mitchell and Hall (2005); Bache *et al.* (2009).

<sup>2</sup>Del Negro *et al.* (2016); Billio *et al.* (2013); Aastveit *et al.* (2018); McAlinn and West (2019).

## 2.2 Econometric Framework

In this section, I begin by describing BPS, followed by a discussion of the synthesis functions. Appendix A.2 provides details on the global-local shrinkage priors, implementation of the factor model combination, and the implementation of BPS, along with an overview of the MCMC algorithm to estimate the density combinations.

### 2.2.1 Bayesian Predictive Synthesis

Bayesian Predictive Synthesis is a method for combining predictive densities.<sup>3</sup> The theory of BPS provides the posterior distribution of the combined density forecast. In other words, given a set of forecasts, for say GDP, BPS provides an expression for the distribution of GDP conditional on those forecasts. This posterior distribution is then estimated with an MCMC routine with two steps. The procedure amounts to estimating a synthesis function, which is used to combine the forecasts, on a set of regressors drawn from the predictive distributions I wish to combine. As pointed out by Aastveit *et al.* (2023), this means BPS can be thought of as a multivariate regression model with generated regressors as predictors.

More formally, the decision-maker  $\mathcal{D}$  is presented with  $h_j(x) \in \mathcal{H}$ , where  $h_j(x)$  are the set of density functions that are elements of the information set  $\mathcal{H}$ , and  $x$  is a (conditional) draw from the forecast distributions. The goal of BPS is to find a distribution of the target variable ( $y$ ) conditional on these densities:  $p(y|\mathcal{H})$ . The agent opinion analysis theory (West and Crosse (1992) and West (1992)), extended to a time series context by McAlinn and West (2019), shows that the posterior has the form:

$$p(y_t|\Phi_t, \mathcal{H}_t) = \int \alpha(y_t|x_t, \Phi_t) \prod_{j=1:J} h_{tj}(x_{tj}) dx_{tj} \quad (2.1)$$

where  $x_t = x_{t,1:J} = (1, x_{t,1}, \dots, x_{t,J})'$ ,  $J$  is the number of experts, and the dimension of  $x_t$  is  $d = J + 1$  to include an intercept that can account for biases.  $\alpha(y_t|x_t, \Phi_t)$  is an arbitrary synthesis function used to combine the expert densities, while  $\Phi_t$  are the

<sup>3</sup>A general description can be found in McAlinn and West (2019) and specific details related to this application can be found in the technical appendix.

## Chapter 2. Combining Large Numbers of Density Predictions with Bayesian Predictive Synthesis

synthesis function parameters. This equation shows how to relate a set of agent forecast distributions to the decision-maker's combined forecast or, in more simple terms, how to combine forecast distributions in a Bayesian fashion. With equation 2.1 in hand, I can write out a Gibbs Sampler with two blocks:

1. Estimate the synthesis function  $\alpha(y_t|x_t, \Phi_t)$  by sampling from  $p(\Phi_{1:t}|y_{1:t}, x_{1:t})$ .
2. Then, draw  $x_{1:t}$  from  $p(x_{1:t}|\Phi_{1:t}, y_{1:t}, \mathcal{H}_{1:t})$ .

As an illustrative example of the MCMC routine, consider the following synthesis function used in McAlinn and West (2019):

$$y_t = x_t\beta_t + \epsilon_t \quad \beta_t = \beta_{t-1} + u_t\epsilon_t \sim \mathcal{N}(0, \sigma_t^2) \quad u_t \sim \mathcal{N}(0, \theta) \quad (2.2)$$

where  $y_t$  is the target variable,  $x_t$  are draws from the forecast distributions (including a vector of ones for the intercept), and dimension  $d = J + 1$ , where  $J$  is the number of experts,  $\beta_t$  are combination weights that vary over time following a random walk with variance  $\theta$ , and  $\epsilon_t$  is an error term with time varying volatility  $\sigma_t^2$ .

The first step in BPS is to estimate equation 2.2, which is a textbook state-space model and can be estimated with standard techniques (Prado and West (2010), Sect 4.5). This is a very flexible specification that can account for biases in the expert's predictions, recalibrate the predictions, and allow for model incompleteness. Applying BPS to different synthesis functions, such as global-local shrinkage priors and factor model combinations, is straightforward. The researcher simply specifies the function and estimates it during the appropriate Gibbs step.

The second step of the MCMC is to draw new forecasts from  $p(x_{1:t}|\Phi_{1:t}, y_{1:t}, \mathcal{H}_{1:t})$ . These  $x_t$  are conditionally independent over time with the following conditionals:

$$p(x_t|\Phi_t, y_t, \mathcal{H}_t) \propto N(y_t|X_t'\beta_t, \epsilon_t) \prod_{j=1:J} h_{tj}(x_{tj}) \quad \text{with} \quad x_t = (1, x_{t1}, \dots, x_{tJ})' \quad (2.3)$$

If the individual expert densities are normal ( $h_j(x_j)$ ), this yields a multivariate normal for  $x_t$  and can be sampled with a Gibbs step using the analytical results from

McAlinn and West (2019). However, the applications in this paper do not have analytical representations. For example, the European SPF elicits histograms from survey respondents. This requires a small adjustment to the algorithm, which is to sample  $x_t$  using a block Metropolis-Hastings step using the aforementioned multivariate normal as a proposal distribution. Details are provided in the appendix. Finally, when creating forecasts (e.g.,  $y_{t+1}$ ), I create “synthetic futures” as in McAlinn and West (2019). That is, for every pass of the MCMC, the synthesis function parameters are iterated forward using the model dynamics and  $x_{t+1}$  are drawn unconditionally from  $h_{jt+1}(x_{jt+1})$ .

### 2.2.2 Global-local Shrinkage Priors

This section discusses the implementation of global-local shrinkage priors in BPS. Global-local shrinkage priors are a common way of introducing shrinkage to Bayesian statistical models. This class of prior includes many commonly used shrinkage priors and gets its name from the two parameters in the prior: one governs shrinkage over all parameters and another governs component-specific shrinkage. More precisely, the prior has the following form:

$$\beta_j \sim \mathcal{N}(0, \kappa\psi_j)$$

where  $\kappa$  is a global shrinkage parameter and  $\psi_j$  is a component-specific parameter. The prior distribution on these individual components determines the shrinkage properties. There are a wide variety of possible choices. In general, a desirable shrinkage profile is horseshoe-shaped, which means there are two modes in the shrinkage density such that coefficients are shrunk to zero or are scarcely changed. For this paper, I use the triple gamma prior (Cadonna *et al.*, 2020) since it has the desirable horseshoe-shaped shrinkage profile and is very flexible, encompassing many other commonly-used priors. Since the triple gamma prior encompasses many priors as special cases, I also consider the horseshoe prior (Carvalho *et al.*, 2010), double gamma (Bitto and Frühwirth-Schnatter, 2019), and Bayesian Lasso (Belmonte *et al.*, 2014). All these priors have fully hierarchical representations, so no tuning of hyperparameters is required. Details are provided in the technical appendix.

I consider time-varying parameter models where shrinkage is imposed on them by rewriting them in the non-centered parameterization (Frühwirth-Schnatter and Wagner, 2010):

$$y_t = x_t\beta + x_t \text{Diag}(\sqrt{\theta_1}, \dots, \sqrt{\theta_d})\tilde{\beta}_t + \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0, \sigma_t^2) \quad (2.4)$$

$$\tilde{\beta}_t = \tilde{\beta}_{t-1} + \tilde{u}_t, \quad \tilde{u}_t \sim \mathcal{N}_J(0, I_J)$$

The non-centered parameterization allows shrinkage on  $\theta$ , which is the variance of  $\beta_t$ , and  $\beta$ , which is the constant component. This means the coefficients can be constant, time-varying, or time-varying with an intercept. In addition, alternating between the centered and non-centered parameterization in the MCMC routine can improve the estimation efficiency (Yu and Meng, 2011; Kastner and Frühwirth-Schnatter, 2014; Kastner *et al.*, 2017). The model in equation 2.4 is estimated by the MCMC described in Cadonna *et al.* (2020) and Bitto and Frühwirth-Schnatter (2019), and a sketch of the algorithm is presented in the appendix.

### 2.2.3 Factor-model-based Combination

The next section discusses how a factor model can be used as a synthesis function in BPS. There are many options for specifying a factor model (Lopes, 2014), but in this paper I follow the classic example from Lopes and West (2004). The Bayesian factor model is a natural choice for synthesis function since macroeconomic forecasts can be highly correlated and it has been successful in many applications with large numbers of predictors.

To see how a factor model can be used as a synthesis function, consider equation 2.2. Simply replace  $x_t$  with  $f_t$  in the observation equation, which is a factor estimated on the draws  $x_t$ . This results in equation 2.5:

$$y_t = F_t' \gamma_t + \epsilon_t \quad \gamma_t = \gamma_{t-1} + u_t \quad x_t = \Lambda f_t + \nu_t \quad (2.5)$$

$$\epsilon_t \sim \mathcal{N}(0, \sigma_t^2) \quad u_t \sim \mathcal{N}(0, \theta) \quad \nu_t \sim \mathcal{N}(0, R) \quad (2.6)$$

where  $F_t = (1, f_t)$  and  $f_t$  is a  $k \times 1$  vector of factors,  $\Lambda$  is a  $J \times k$  vector of loadings, and  $R$  is a diagonal covariance matrix with elements  $\sigma_{\nu J}^2$ . Additionally,  $\gamma_t$  is a  $k + 1$  vector. In order to derive combination weights, I need to identify the factors. This is done by using the restriction  $f_t' f_t = I_J$  and restricting the first  $k$  rows of the loadings matrix to be the lower diagonal and positive elements on the main diagonal. This is a common identification scheme used to fix indeterminacy in the estimation of the factors.

MCMC estimation is straightforward since the loadings can be estimated by linear regression and the factors can be drawn from a conditional normal distribution. The  $x_t$  are standardized using the mean and standard deviation estimated from the marginal distribution of each agent. There is a small complication introduced by the factor model when drawing  $x_t$ . This is because I need to evaluate  $p(y_t | x_t \gamma_t, \epsilon_t)$  during the MH step, but equation 2.5 is specified in terms of  $f_t$ . However, the model can be reparameterized in terms of  $x_t$  and  $x_t | y_t, \Phi_t$  and sampled using the standard technique.

## 2.3 Forecasting Environment

I empirically compare the sparse and dense combination approaches in two settings. The first exercise is nowcasting Canadian real GDP in pseudo real-time using a large set of models. This is a good application for two reasons. First, it is a common scenario considered by policy-makers and researchers at Central Banks.<sup>4</sup> And, second, a nowcasting cycle allows for a comprehensive assessment of performance since it involves multiple forecast horizons, a diverse set of models, and large datasets with a mix of hard and soft indicators. Overall, a nowcasting application provides a realistic and challenging environment for the various synthesis functions. The second exercise is forecasting Euro Area real GDP using the Survey of Professional Forecasters, which is a more standard setting than nowcasting in Canada.<sup>5</sup> Using the SPF is a good check on the Canadian results, which may be affected by idiosyncratic factors. These two environments are very different: not only are they in different regions, but the types of forecasts provided are also different. The nowcasting exercise uses model-

---

<sup>4</sup>Examples are Aastveit *et al.* (2011), Chernis and Webley (2022), and Knotek and Zaman (2023).

<sup>5</sup>For example, Diebold *et al.* (2023) and Conflitti *et al.* (2015) consider the European SPF.

based predictions from four different model classes. In contrast, the forecasting exercise features mostly judgemental forecasts (ECB, 2019) that are provided as histograms. Since these two applications cover two regions, have different forecast horizons, include model-based and survey-based predictions, and have an evaluation sample that covers the Great Financial Crisis, Euro Area Crisis, and COVID-19 pandemic, they should allow for a comprehensive and reliable assessment of the various synthesis functions.

### 2.3.1 Details on the Model-based Nowcasting Exercise

The first application uses density predictions produced in Chernis and Webley (2022), which builds on Chernis and Sekkel (2018), as inputs into BPS. The models are standard implementations of nowcasting models used at Central Banks. They include leading indicator models (or ARX), mixed data sampling models (MIDAS), Bayesian vector autoregression, and dynamic factor models, totalling 98 models (see figure 2.1). Predictions are made using a medium-sized dataset of 35 indicators that is constructed by choosing variables that are followed by the market and, in many cases, reported on Statistics Canada’s official release bulletin “The Daily.” It includes 24 domestic indicators, seven US or international indicators, and four financial variables. The reader can consult these papers and references within for detailed results and descriptions of the models and dataset. Pseudo real-time forecasts are produced from 2000 to 2021 and real-time predictions from 2013 to 2019. In this paper, I use the pseudo real-time forecasts with a five-year expanding estimation window; the training sample covers 2000–2005, with an evaluation window from 2005Q1 to 2021Q1.

Figure 2.1: Model List

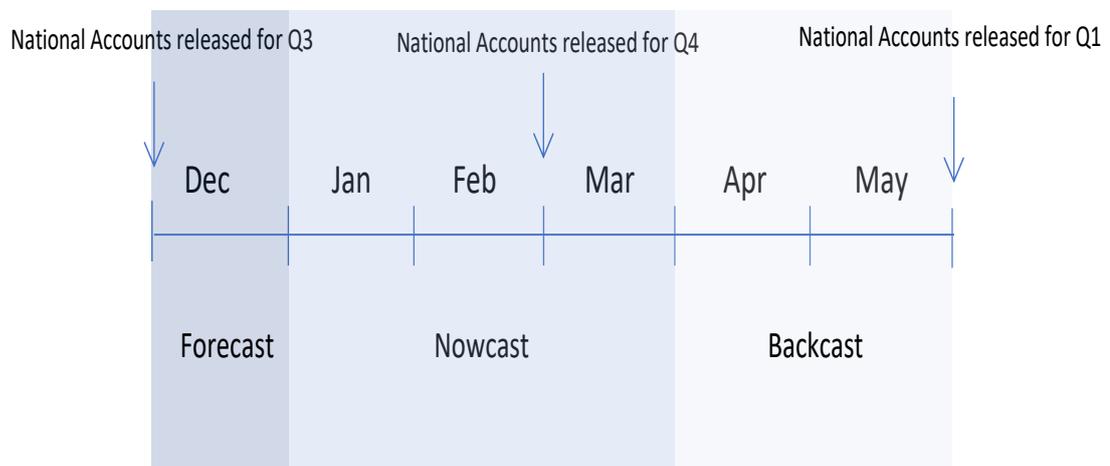
Model Class	Density Method	Number of Models	Estimation Window
ARX	Block Wild Bootstrap	48	Rolling/Expanding
MIDAS	Block Wild Bootstrap	24	Expanding
BVAR	Bayesian Methods	22	Rolling/Expanding
DFM	Bayesian Methods	4	Rolling/Expanding

In a nowcasting exercise, the timing of the forecast cycle can be quite important. Figure 2.2 illustrates the timing of releases throughout the six-month forecast cycle

## Chapter 2. Combining Large Numbers of Density Predictions with Bayesian Predictive Synthesis

starting in December after the release of the Q3 National Accounts data targeting the Q1 figures for the upcoming year. Forecasts are produced 12 times over the six months, representing a prediction roughly every two weeks, and is designed to replicate the forecast cycle faced by a practitioner. The cycle starts in December, when the analyst is forecasting the Q1 figures. Throughout Q1, the analyst is in the nowcast phase. From the April to May National Accounts data release, the analyst is backcasting the Q1 figures while awaiting publication of the official figures.

Figure 2.2: Overview of Forecast Cycle



A peculiarity of Canadian nowcasting is that there is a monthly GDP figure available two months after the reference period. This data is different from the National Accounts figures since monthly GDP is at a by-industry basis compared to the expenditure approach of the National Accounts. There can be differences between the figures of as much as a percentage point. This means monthly GDP is an important predictor for quarterly GDP, but not a perfect predictor. The consequences of including this predictor in the dataset is that once it is available, there is a large improvement in the accuracy of the prediction and other variables become less important (Chernis and Sekkel, 2017).

### 2.3.2 Details on the Survey of Professional Forecasters

The survey forecast application uses density forecasts from the European Central Bank’s Survey of Professional Forecasters. A full description is available in García (2003). The quarterly survey began in 1999 and is the longest running Euro-area survey of macroeconomic expectations. The survey elicits probability and point forecasts on inflation and GDP growth at various horizons (I use the one-year-ahead expectation for year-over-year GDP growth). On average, there are 50 responses a quarter from a panel of over 100 participants. Because of the time series length and panel characteristics, the survey is often used to study density forecast combinations as seen in Diebold *et al.* (2023) and Conflitti *et al.* (2015).

Several attributes of the survey merit discussion. Survey respondents are provided with fixed ranges for which they provide probabilities. For example, in 1999Q1, they were provided with 10 bins, the first starting with less than 0 percent and increasing by 50 basis point intervals to 4 percent growth or above. A few issues arise here. First, the bins change over time to address unexpected developments (such as the COVID-19 shock) and the open intervals. The bins changing over time are not an issue for the model since I convert the forecasts to pdfs over a fine grid of 750 points. This results in a pdf resembling a histogram, and adding more bins just adds more rectangles to the pdf. For the open bins, I distribute the assigned probability, if any, from the start of the bin plus or minus two standard deviations of GDP growth, estimated using the vintage available at the time of the forecast.

Another issue is that forecasters can join and leave the panel at any time. This means there are often missing forecasts and the panel size can change over time. This paper takes two approaches to deal with survey entry and exit in an effort to avoid the results being influenced by these choices. First, I construct a “wide” dataset with the goal of including as many forecasters as possible, corresponding to the approach taken in Conflitti *et al.* (2015). Since there are a large number of missing forecasts, I drop forecasters with fewer than five forecasts in a five-year period. This five-year period is also the rolling estimation window I use for the model. After dropping forecasters for each five-year period, the remaining unbalanced panel has about 35 respondents

each quarter. One consequence of the changing panel composition is that examining the online weights is not meaningful because there are different forecasters at each point in time. Next, missing observations in the panel are imputed. Deviating from Conflitti *et al.* (2015), these missing distributions are filled in with a normal distribution corresponding to the marginal distribution of GDP estimated in real time.<sup>6</sup> Overall, this is a very challenging prediction exercise since there are large amounts of missing data, a wide panel, and a short time series to train the algorithm.

Second, I construct a “tall” dataset that aims to build the longest consistent panel possible. Following Diebold *et al.* (2023), I drop forecasters who have not responded for five consecutive quarters. This results in a panel of 14 forecasters with minimal missing data. Any missing data is imputed with a normal distribution corresponding to the unconditional distribution of GDP estimated in real-time. Despite having half as many experts as the “wide” dataset relative to the length of the panel, this is still a wide dataset. However, the prediction exercise is easier than the “wide” dataset since there is much less missing data being imputed and a longer time series to train the algorithm.

Once the data set is assembled, the first estimation window is 1999Q3 to 2004Q2, and the evaluation window is 2005Q2 to 2020Q4. The forecast combination is estimated with a five-year rolling window for the “wide” dataset and an expanding window for the “tall” dataset. This is a full real-time exercise with the models estimated on the vintage available to the forecasters and evaluated against the most recently available vintage of GDP.

## 2.4 Results

This section discusses the main findings of this paper: 1) sparse combination techniques often perform better, and 2) the simpler constant parameter combinations usually perform better. First, there are a few details to dispense with. I use the CRPS as an accuracy metric since it is less sensitive to outliers relative to the log score and thus prevents extreme events from dominating performance during “normal” times. Detailed

---

<sup>6</sup>Replacing missing forecast distributions with a uniform distribution, as in Conflitti *et al.* (2015), does not qualitatively change the results.

results for the nowcasting exercise are reported in table A.1, and table A.3 shows results for the survey forecast application. The tables also include as a benchmark the dynamic linear model (DLM) in equation 2.2, which is used in McAlinn and West (2019). Overall, the synthesis functions introduced in this paper are competitive with the benchmark, suggesting that the performance of the new synthesis functions are reasonable. Finally, in the appendix I present results suggesting that forecasts from both synthesis functions are well calibrated using the test from Knüppel (2015).

### 2.4.1 Predictive Accuracy: Global-local Shrinkage Priors and Factor Model Combinations

In general, the shrinkage priors have lower average CRPS across both nowcasting and forecasting exercises. Figure 2.3 shows the results from the nowcasting exercise (panel (a)) and for the survey forecast application (panel (b)). For clarity of exposition, the results from factor models are shaded red and the shrinkage priors are shaded blue. Most of the time the shrinkage priors perform better, and this pattern is apparent in both the nowcasting and the SPF application across both the tall and wide datasets. Depending on the application and forecast horizon, improvements can be as high as 30 percent in the nowcasting exercise and 20 percent in the SPF application. There are some exceptions, which I examine throughout the remainder of this section.

To better understand the forecasting performance over time, I examine the cumulative CRPS difference. This is particularly useful for highlighting episodes that may have undue influence on average forecast accuracy and help explain the above findings. For brevity, I focus my analysis on comparing the best-performing models from each class of synthesis function: the constant parameter triple gamma prior and the factor model combination with two factors. Figure 2.5(a) shows results for the nowcasting application and panel 2.5(b) for the survey forecast exercise. Despite being for different countries and different forecast horizons, there are similarities in the results. In both applications, the triple gamma prior performs slightly worse at the beginning of the sample. However, as the GFC arrives, the triple gamma prior begins to perform better, signified by positive values in the figures. For most of the post-GFC period, the

Chapter 2. Combining Large Numbers of Density Predictions with Bayesian Predictive Synthesis

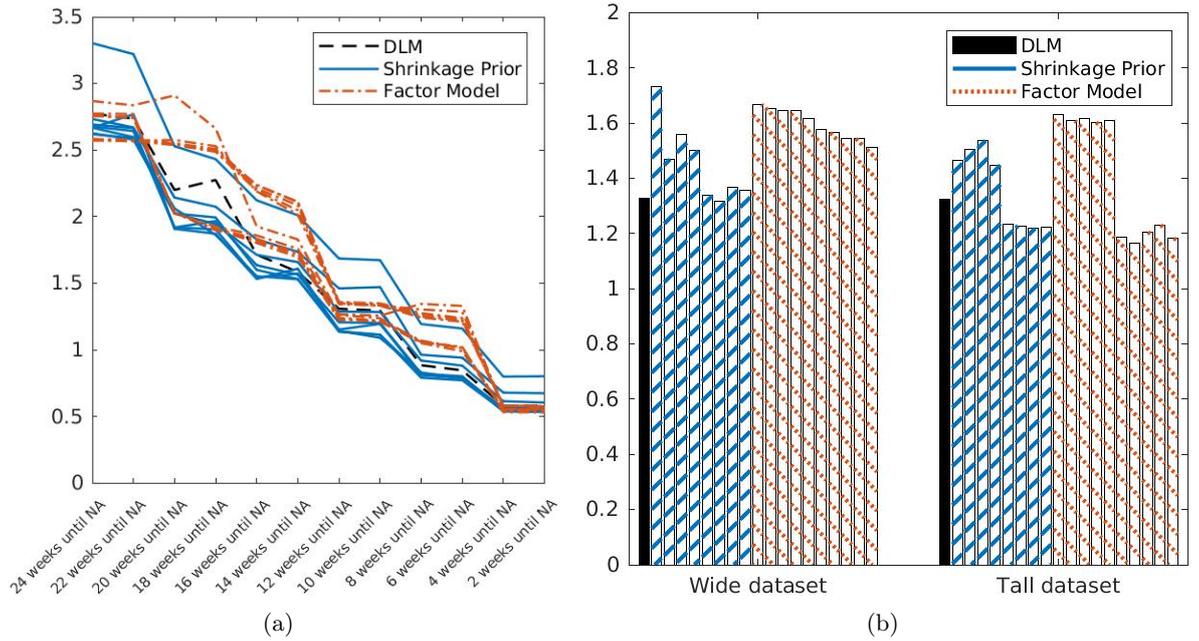


Figure 2.3: Comparison of Shrinkage Priors and Factor Models

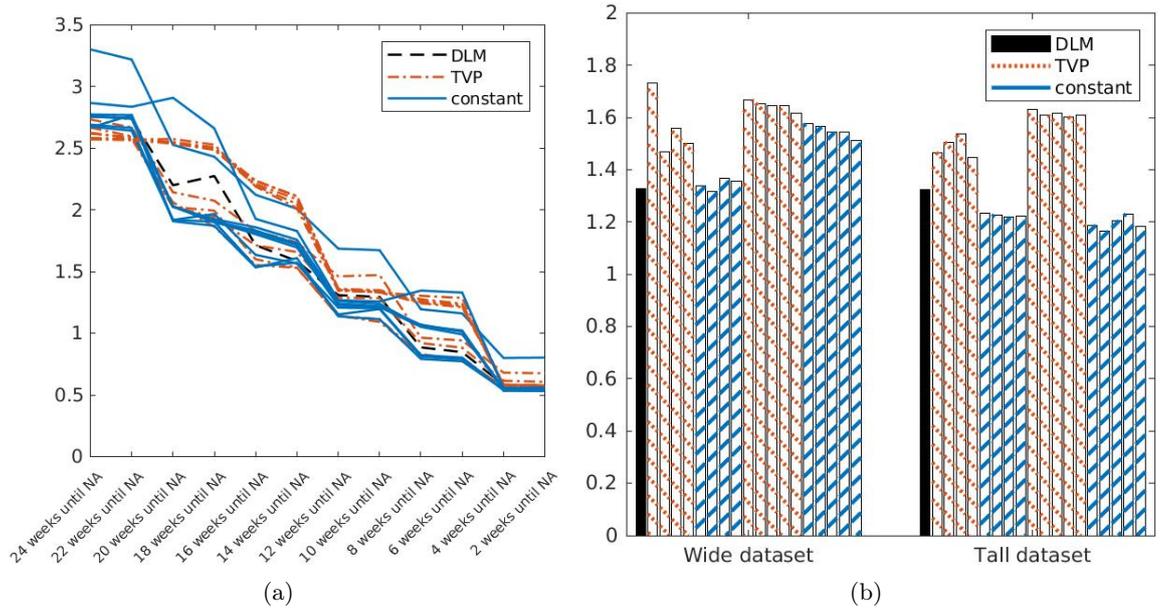


Figure 2.4: Comparison of Time-varying and Constant Parameter Specification

triple gamma prior continues to improve upon the factor model combination. During the COVID period, the factor model starts to perform better in both the SPF tall data

## Chapter 2. Combining Large Numbers of Density Predictions with Bayesian Predictive Synthesis

application and the longer prediction horizons in the nowcasting application. Inspecting the forecast densities over time can help explain these results.

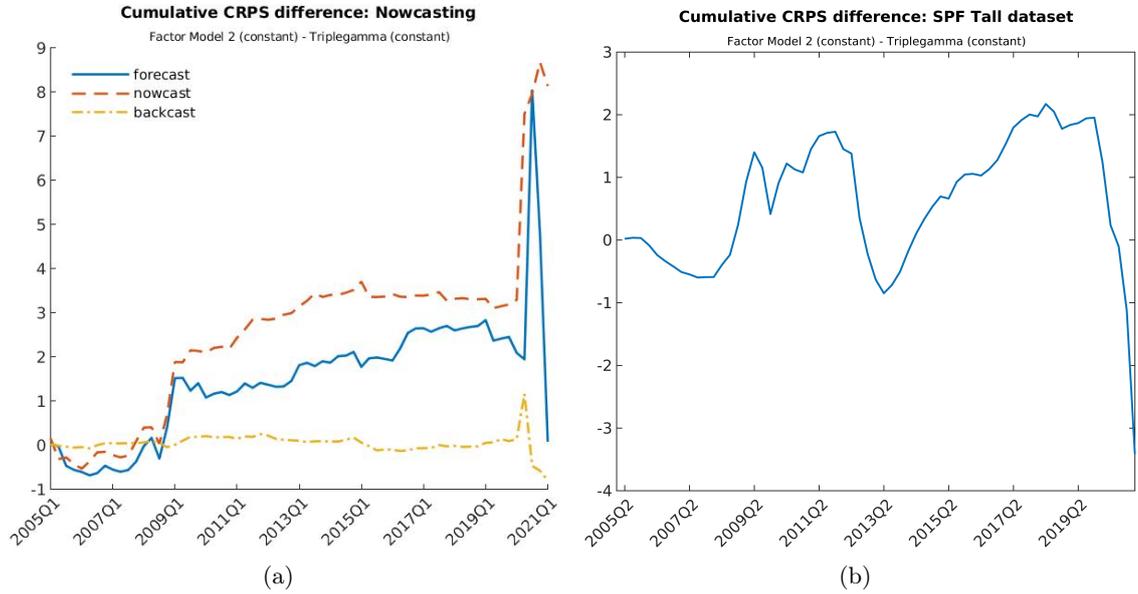


Figure 2.5: Cumulative CRPS Difference

Figure 2.6 shows predictive densities for the nowcasting application<sup>7</sup> and figure 2.8 shows the difference between their forecast probability distributions. Red shading indicates higher probability assigned to a region by the triple gamma prior, and blue indicates more probability assigned to a region by the two factor model. Similar charts for the SPF application are shown in figures 2.7 and 2.9.

Remarkably, examining the forecast densities shows that there are similar patterns across the forecasting and nowcasting applications. In both applications, GDP out-turns often occur in the red areas where the triple gamma combination put higher probability.<sup>8</sup> Additionally, the triple gamma prior has lower variance predictions than the factor models, which contributes to their better performance. This is most obvious post-GFC in the SPF application where the shrinkage priors produce substantially lower variance forecasts using both the wide and tall datasets and the forecast periods of the nowcasting application. The more precise predictions of triple gamma prior results in

<sup>7</sup>COVID-19 is excluded because the volatility in GDP growth makes the chart illegible.

<sup>8</sup>This result is corroborated by a quantile score decomposition of the CRPS, which shows better performance across the entire distribution. Results are available upon request.

Figure 2.6: Nowcasting Application Predictive Densities

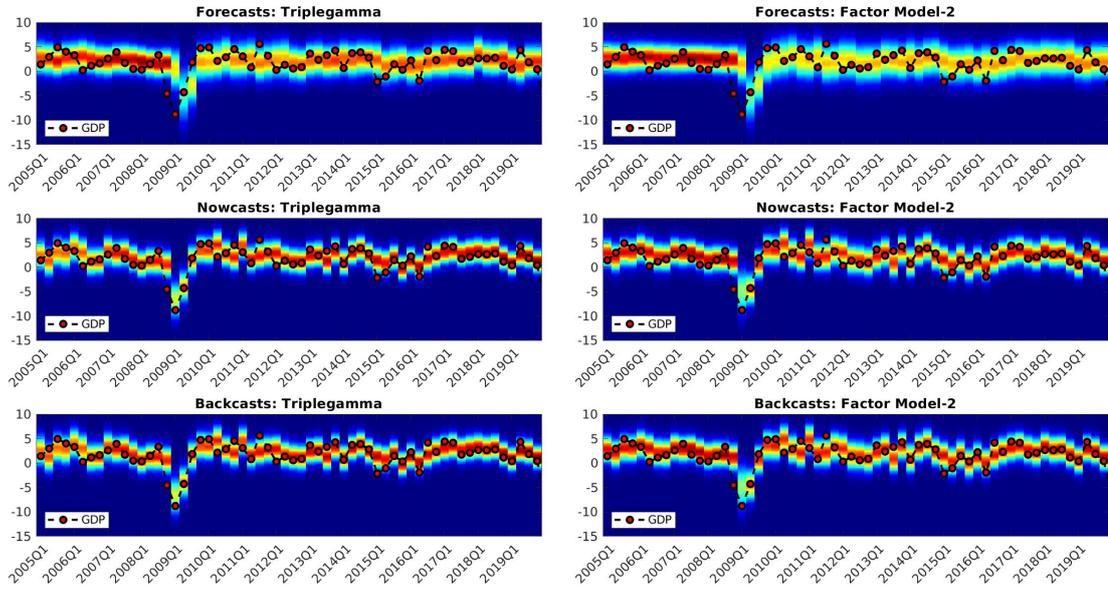
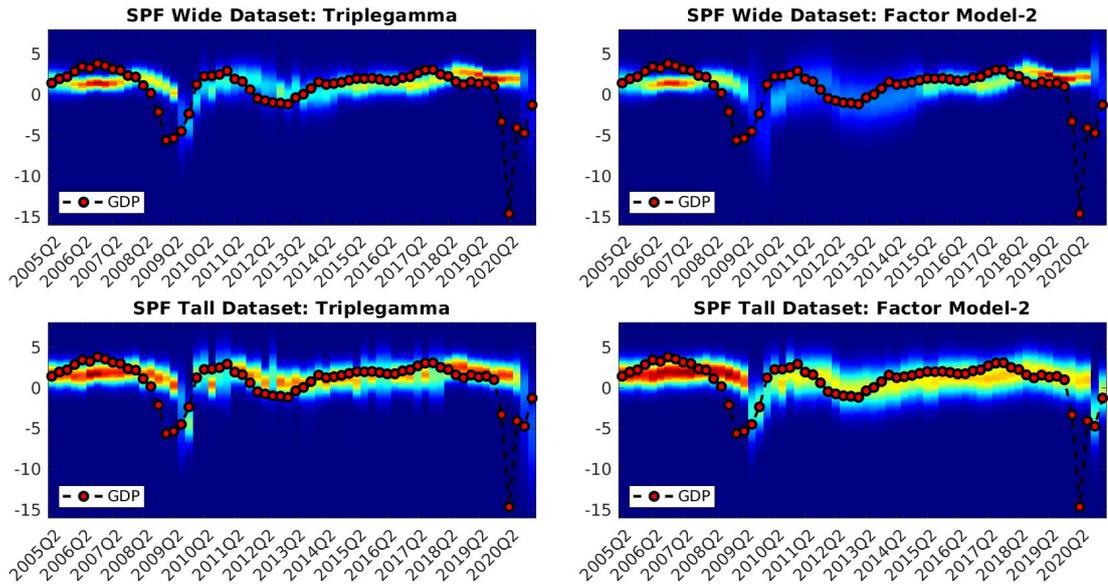


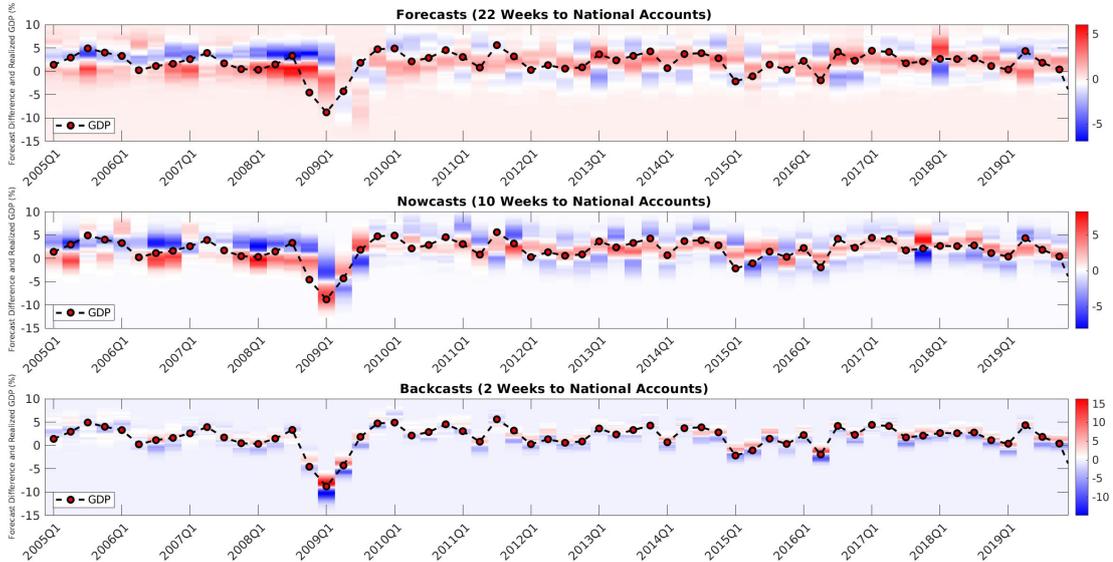
Figure 2.7: SPF Application Predictive Densities



systematically better forecasts during normal times. This can be seen in figure 2.9 by noticing the large number of out-turns in the red shaded regions, which signify the shrinkage prior places relatively higher probability on the out-turn. However, there is an important caveat to this result. Figure 2.5(b) shows that during the Euro Area crisis, the factor model improves over the triple gamma approach. This can be seen in

## Chapter 2. Combining Large Numbers of Density Predictions with Bayesian Predictive Synthesis

Figure 2.8: Nowcasting Application Difference Between Triplegamma and Factor Model-2 in Predictive Densities with GDP Realizations



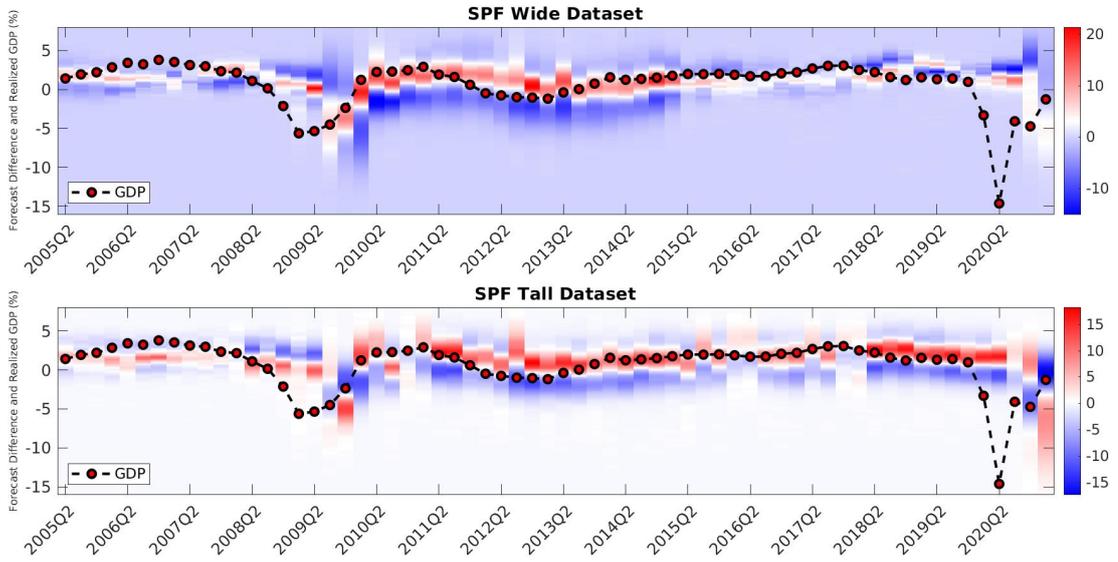
Notes Red values in the heat map indicate the triple gamma prior adds more probability to the bin relative to the factor model. Blue shading signifies the factor model adds more probability to a region than the triple gamma prior.

figure 2.9 where the GDP out-turns are in the blue area, signifying higher mass put on that region by the factor model. It turns out that this is because one of the forecasters, which has a weight of around 75 percent, drops out from the sample for a few quarters and is replaced with the unconditional distribution of GDP. This results in poor forecast performance and, when combined with overconfidence, is quite punishing. The factor model approach has more egalitarian weights, so this is less of a problem. This serves as a practical lesson that placing significant weight on an individual expert has risks.

An event deserving special scrutiny is the COVID-19 pandemic. Not surprisingly, it has a significant impact on the results and is the reason why the factor model combination is occasionally competitive. For the nowcasting application, this is most apparent at the forecast horizon and to a lesser extent when backcasting. The triple gamma captures the declines in 2020Q1 and 2020Q2 more accurately, but it misses the sharp and immediate rebound in 2020Q3. This can be observed in figure 2.5(a), where the blue line approaches the y-axis at the end of sample, signifying that on average the two

## Chapter 2. Combining Large Numbers of Density Predictions with Bayesian Predictive Synthesis

Figure 2.9: SPF Application: Difference Between Triple Gamma and Factor Model-2 in Predictive Densities with GDP Realizations



Notes Red values in the heat map indicate the triple gamma prior adds more probability to the bin relative to the factor model. Blue shading signifies the factor model adds more probability to a region than the triple gamma prior.

methods perform similarly. There is a similar result for the SPF application using the tall dataset. During the pandemic, the factor models perform so well that they catch up to and slightly exceed the average performance of the shrinkage priors. Figure 2.7 shows that in 2020Q1 and 2020Q2, this is not so much due to the factor model providing a significantly better forecast. The higher variance of the factor model combination means they are punished less for inaccurate predictions. The story is different in 2020Q4 and 2021Q1, where the factor model is more accurate. This can be seen quite clearly in figure 2.7, where the factor model not only puts a large amount of mass around the out-turns, but also relatively more than the triple gamma (figure 2.9). The reason for this is that triple gamma prior puts close to 70 percent of the weight on a single model, which happens to provide a very poor forecast. Again, this highlights the risks of using sparse weights. While the triple gamma synthesis function systematically outperforms the factor models most of the time, it can be risky to put a large amount of weight on a single expert.

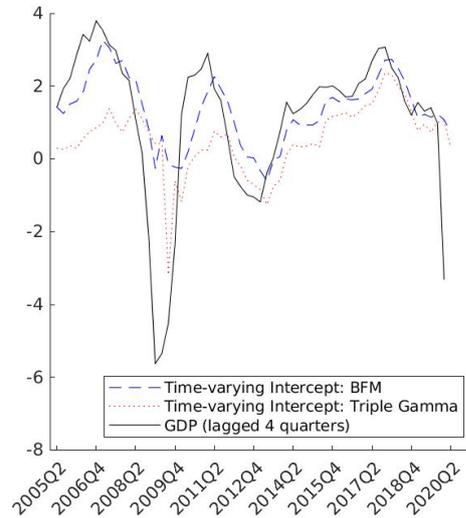
### 2.4.2 Predictive Accuracy: Time-varying and Constant Parameter Combinations

Another finding is that constant parameter combinations generally have a lower CRPS than their time-varying counterparts. Figure 2.4 shows the CRPS for the nowcasting (panel (a)) and the SPF application (panel (b)), which shades the time-varying parameter combinations red and their constant parameter counterparts blue. There can be significant gains for choosing the more parsimonious constant parameter specification. In the nowcasting application, there are gains of up to 20 percent between constant and time-varying factor model combination specifications. In the survey forecasting exercise, performance gains can be up to 25 percent for both shrinkage and factor model combinations, and improvements are seen in both datasets for both classes of synthesis functions.

The most dramatic performance increases, seen in the survey forecast application, are explained by the time-varying parameter combinations reducing to a time-varying mean model with little weight on the individual experts. Figure 2.10 shows the in-sample time-varying intercept for the triple-gamma prior and the one-factor combination approach overlaid with the four-quarter lagged Euro Area GDP figures. It is apparent that the intercept matches the GDP figures very closely, suggesting that it may be overfitting. Additionally, inspection of the weights for each of the time-varying combination methods reveals that the weight put on individual experts is quite small (lower panel in figure 2.11). In contrast, the upper panel of figure 2.11 shows that the sum of weights from constant parameter specifications is much closer to 1. Taken together, it is evident that poor performance of the time-varying parameter models is because the forecast is driven by the time-varying intercept while ignoring useful information contained in the expert densities. On the other hand, the constant parameter specifications, which lack the flexibility of a time-varying intercept, place more weight on the experts. This finding is in contrast to other studies (Aastveit *et al.* (2023)), which find that a time-varying intercept in BPS can be extremely useful. This is likely due to differences in the applications—the aforementioned paper forecasts oil prices, which have large and persistent movements in price that make a time-varying intercept useful. In contrast,

Euro Area and Canadian real GDP have much smaller movements in their growth rates over the twenty-year period in question.

Figure 2.10: Time-varying Predictive of Mean of BPS Intercepts for SPF Tall Dataset



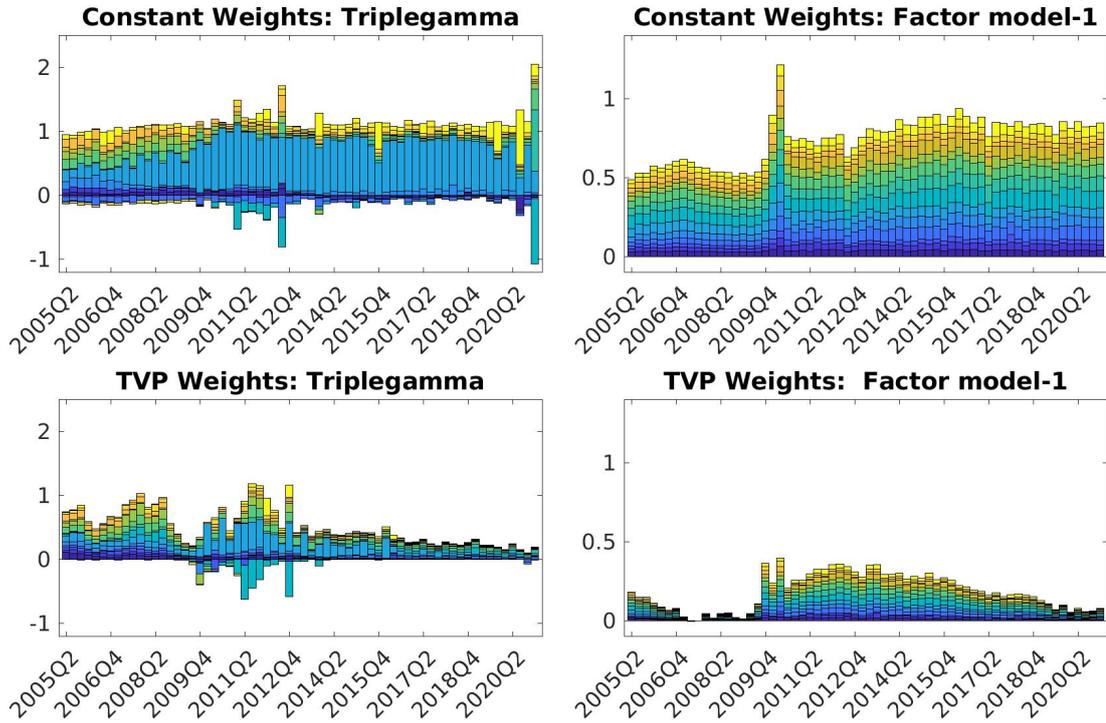
### 2.4.3 Examining the Combination Weights

It is instructive to examine the weights in figure 2.11 to gain some intuition of the implications of synthesis function choice. Let us start with the weights from the triple gamma prior in the top left panel.

First, the combination method puts significant weight on a single expert, a handful of other forecasts, and close to zero weight on the rest.<sup>9</sup> This prior implies the decision-maker should mostly listen to a few trusted experts, but not completely ignore the herd. Additionally, a few of the experts have negative weights. This reflects the very flexible specification that allows the weights to adjust for biases. This is similar to portfolio optimization where the optimal portfolio involves short selling an asset as a hedge. Put in terms of BPS, the decision-maker hedges against the high weight on a given expert by “short-selling” a similar correlated forecast.

<sup>9</sup>The triple gamma appears to be good at picking up weak signals in the data and not shrinking experts to zero weight. Sparsifying the weights using signal adaptive variable selector (Ray and Bhattacharya (2018)) results in worse forecasting performance, suggesting the non-zeros weights are not numerical artifacts.

Figure 2.11: Sequentially Estimated Mean Combination Weights



Second, examination of the right panel indicates that the one factor model has weights that are spread more evenly over experts (but not equally), meaning the combination is closer to consensus weights.<sup>10</sup> I use the term “consensus weights” since a factor model extracts the common variance across experts or, in some sense, what the experts can agree upon. There is an important difference between this weighting scheme and equal weights since the former removes idiosyncratic differences across experts and the latter includes all the experts equally. This synthesis function implies the decision-maker should follow a consensus-based approach to processing forecasts, and the approach is quite different from shrinkage priors where the decision-maker focuses on a small subset of experts. The results above suggest that sparse weights are preferable to consensus weights—a decision-maker should not follow the herd, but instead focus on a smaller set of experts.

<sup>10</sup>Adding more factors allows experts to have more weight but does not change the pattern of dense weights or the interpretation of consensus-based weights.

## 2.5 Conclusion

In this paper, I investigate different approaches for combining large numbers of density predictions within the framework of Bayesian Predictive Synthesis. This is an important issue since many practical applications can involve large numbers of forecasts, such as nowcasting systems or combining survey forecasts. I use two common approaches in economics to deal with large datasets: global-local shrinkage priors and factor modeling. In particular, I use the newly developed triple gamma prior, and the priors it encompasses, along with a novel factor modeling approach to density combinations.

I test the approaches using two very different applications: a model-based nowcasting exercise on Canadian real GDP, and forecasting Euro Area real GDP growth using distributions from the Survey of Professional Forecasters. These two applications cover two regions, have different forecast horizons, include model-based and survey-based predictions, and the evaluation sample covers the Great Financial Crisis, Euro Area Crisis, and COVID-19 pandemic, allowing for a comprehensive assessment of the various synthesis functions. First, I find that constant parameter specifications tend to perform better than their time-varying counterparts. This shows that in applications with little structural change, relatively short samples, and a large cross-section of models, a more parsimonious model is preferable. This is an important finding as recently developed combination schemes tend to utilize time-varying parameter specifications. Second, and more importantly, I find that shrinkage approaches generally outperform factor-model-based combinations. With the exception of the Bayesian lasso, the shrinkage priors all perform well in terms of a low average CRPS.

It is interesting to note that the two synthesis functions imply very different weighting structures. The sparse weighting scheme of shrinkage priors implies that decision-makers should give considerable weight to a smaller set of experts. This, however, carries the risk of “putting all your eggs in one basket,” which at times adversely affects the performance of the sparse combinations. In contrast, the factor-model-based combination implies a dense weighting scheme, which produces a “consensus” forecast. Overall, my results suggest that focusing on a parsimonious combination that considers

## Chapter 2. Combining Large Numbers of Density Predictions with Bayesian Predictive Synthesis

a smaller set of accurate experts is preferable to following the herd.

## Chapter 3

# Predictive Density Combination Using a Tree-Based Synthesis Function

### 3.1 Introduction

It is commonplace when forecasting macroeconomic variables, such as output growth or inflation, to consider a large number of competing predictive densities. These density forecasts might come from different reduced-form or structural models and/or be subjective and come from surveys. How to combine these densities is an open question being addressed by a growing literature.<sup>1</sup> The literature concludes that combined density forecasts tend to be more accurate and more robust than single-model approaches that ignore model uncertainty; for a review, see Aastveit *et al.* (2019). One issue is that traditional forecast combination techniques are often linear and do not exploit information besides the forecasts and the target variable. Contrast this with a policymaker who combines forecasts nonlinearly and uses external information, such as on the current state of the economy or financial conditions, to help determine how much weight

---

<sup>1</sup>See, among many others, Mitchell and Hall (2005); Wallis (2005); Hall and Mitchell (2007); Geweke and Amisano (2011); Koop and Korobilis (2012); Billio *et al.* (2013); Aastveit *et al.* (2014); Conflitti *et al.* (2015); Chernis and Webley (2022); Knotek and Zaman (2023); Aastveit *et al.* (2023); Čapek *et al.* (2023); Diebold *et al.* (2023).

to attach to the different forecasts. We propose a novel technique that mimics this practice. Our approach combines density forecasts nonparametrically while allowing the combination weights to be determined by information that may be external to the forecasting models.

Key to our approach is Bayesian predictive synthesis (BPS). It has emerged, as extended into a time-series context by McAlinn and West (2019), as a general method of density forecast combination with a strong theoretical basis. BPS draws on an earlier Bayesian literature on agent or expert opinion analysis (West, 1992) and provides a formal and theoretically justified method for pooling densities. It can be shown to nest many previous approaches (see, for example, Section 2.2 of McAlinn and West, 2019) and has been used successfully in various applications in economics, such as McAlinn *et al.* (2020), Chernis (2023), and Aastveit *et al.* (2023). In this paper we develop density forecast combination strategies within the BPS framework.

In existing implementations of BPS, the so-called synthesis function, which determines the weight attached to each density, needs to be specified parametrically. Common choices, as made in the aforementioned papers, are to assume that the synthesis function takes the form of a dynamic linear regression, with parameters that are allowed to change over time typically as random walk processes. This specification of the synthesis function thus allows the weights on competing density forecasts to evolve over time as linear Gaussian random walks. But such an assumption may or may not be valid. Misspecification occurs if the weights depend on other factors or if they follow a different law of motion than a random walk.

These considerations motivate the present paper. BPS has theoretically rigorous foundations, but the manner in which it has been implemented in practice risks misspecification due to the adoption of particular and untested parametric assumptions. We therefore propose to use flexible nonparametric techniques to specify the synthesis function. Specifically, we use regression trees. In conventional (single-model) forecasting applications, tree-based models of the conditional mean have proven highly successful (see, for example, Clark *et al.*, 2023; Huber and Rossini, 2022; Huber *et al.*, 2023). A small number of other papers have used nonparametric techniques to combine predic-

tive densities (for example, Jin *et al.*, 2022; Bassetti *et al.*, 2018, 2023). However, unlike our proposed method, these other papers neither use regression trees nor fit explicitly within the formal BPS framework.

While regression trees have become a popular way to estimate nonparametric regressions, here we propose to use them differently. Similarly to Goulet Coulombe (2024), Deshpande *et al.* (2020), and Hauzenberger *et al.* (2023), who provide a nonparametric treatment to the parameters rather than the variables in single-equation and VAR models, we model the coefficients in the synthesis function with regression trees (RT). Accordingly, we label our version of BPS “BPS-RT.” The synthesis function remains linear in the parameters, which, as we will demonstrate, aids in interpretation. Use of regression-tree methods requires the choice of covariates, which we call “weight modifiers.” These weight modifiers help determine the weights attached to the competing density forecasts. Conventional BPS does not make use of weight modifiers, given that the weights are typically assumed to follow random walks. Thus, in popular implementations of BPS, any relevant information in the form of additional covariates is neglected.<sup>2</sup> But decision makers, when combining competing density forecasts, may wish to condition their forecasts on such “outside” information. For example, they may wish to let the weights on the different forecasting models vary with the state of the economy or vary as a function of the features of each forecast density. Our tree-based specification for the synthesis function is able to condition on both “global” (that is, information not associated with a particular forecaster) and “local” (that is, information associated with a given forecaster) variables when determining the weights. In our tree-based synthesis function, the weights on each density forecast are dynamically determined via a sequence of decision rules. BPS-RT allows the decision maker to combine predictive densities in a highly flexible way and to distill optimally all relevant information contained in the predictive densities and weight modifiers. The fact that the synthesis

---

<sup>2</sup>Notable recent exceptions are Oelrich *et al.* (2023), who, following Li *et al.* (2023), let the weights in linear density forecast combinations depend on (potentially time-varying) exogenous variables. As Oelrich *et al.* (2023) explain, such linear pools are one specific instance of BPS. Letting the combination weights in linear pools change over time according to these “pooling variables,” as in the more general (nonlinear) BPS framework that we consider, can offer more flexibility than assuming that the combination weights follow an assumed autoregressive process; cf. Del Negro *et al.* (2016).

function remains conditionally linear in the parameters helps the decision maker interpret the combined density and better understand the role each individual density is playing in the combination. We will show how BPS-RT can be used to understand the role of model incompleteness, agent clustering, and the time-varying importance of the different weight modifiers.

The next section of the paper introduces and motivates BPS in theory and then discusses how it has been implemented in the existing literature. It then proposes our generalization, BPS-RT, and explores its properties. Section 3 demonstrates the utility of BPS-RT by undertaking two forecasting applications. The first application takes the individual forecaster density forecasts from the European Central Bank Survey of Professional Forecasters (ECB SPF) and combines them. The second application forecasts US inflation using a commonly used large set of indicators. The predictive densities that are synthesized are produced by regression models using the different indicators. We find that BPS-RT produces well-calibrated and accurate forecasts. Notably, we find that single-tree models perform best, in contrast to standard recommendations when using regression trees. This suggests that a relatively parsimonious weight scheme with few changes in weights is supported by the data. The superior performance of BPS-RT stems from its better ability to explain periods of volatility, such as the global financial crisis that affected euro area GDP growth and the post-COVID inflation period in the US. Zooming in on the best-performing BPS-RT specification in the US inflation application, we show how the combination forecasts from BPS-RT can be interpreted. BPS-RT can be used to understand the role of model incompleteness, agent (forecast) clustering, and the time-varying importance of the different weight modifiers. We find little model set incompleteness during the post-COVID inflation period, suggesting that BPS-RT's success comes from its ability to successfully forecast inflation using the underlying models with changes in the combination weights driven by a time trend. This contrasts with the earlier period of lower inflation, when business cycle indicators are shown to be more important. Section 4 concludes. Appendix B.1 provides details on Bayesian inference of BPS-RT, and Appendix B.2 provides additional empirical results, as referenced in the main paper.

## 3.2 Bayesian Predictive Synthesis with Regression Trees

In Section 3.2.1, we provide some background on BPS, distinguishing between BPS in theory and its use in practice in extant empirical applications. Then, in Section 3.2.2, we explain how regression trees can be used to provide a more flexible way of operationalizing BPS.

### 3.2.1 Bayesian Predictive Synthesis

#### BPS in Theory

BPS is a foundational theoretically coherent Bayesian method for combining predictive densities.<sup>3</sup> The theory of BPS provides a pooled predictive distribution for the variable being forecast (say, GDP growth) given a set of individual density forecasts. Operationally, this pooled predictive distribution is produced using Markov chain Monte Carlo (MCMC) methods involving two steps. In the first step, draws are taken from the individual predictive densities for GDP growth. These draws are then, in effect, treated in a second step as explanatory variables in a time-series model where the dependent variable is the outcomes for GDP growth. This time-series model amounts to the synthesis function. Standard choices for this function are typically based on linearity, either simply a constant linear relationship or a dynamic relationship where the linear coefficients evolve over time according to a random walk. As pointed out by Aastveit *et al.* (2023), this means that BPS can be thought of as a multivariate regression relating the target variable (GDP growth) to the forecasts for GDP growth, which are treated as generated regressors. We make use of this generated regressor interpretation below.

More formally, at time  $t$  a decision maker  $\mathcal{D}$  is confronted with  $h$ -step-ahead forecast densities for variable  $y_{\tau+h}$  produced by  $J$  different agents, experts, or models, where  $\tau$  ranges from 1 to  $t$ . At each forecast origin,  $\tau$ , we label these predictive densities  $\{\pi_{j\tau}(y_{\tau+h})\}_{j=1}^J$ . These densities, available from time 1 through  $t$ , form the information set  $\mathcal{H}_t$  of  $\mathcal{D}$  and can, in principle, be of any distributional form.  $\mathcal{D}$  then forms an incomplete joint prior  $p(y_{t+h}, \mathcal{H}_t) = p(y_{t+h}) \times \mathbb{E}\left(\prod_j \pi_{jt}(x_{jt+h|t})\right)$  with

---

<sup>3</sup>For a general description of BPS, see McAlinn and West (2019); specific implementation details related to our applications are discussed below and in Appendix A.

$\mathbf{x}_{t+h|t} = (x_{1t+h|t}, \dots, x_{Jt+h|t})'$  denoting latent agent states (that is, draws from the agent-specific predictive densities). These agents' forecasts target  $t + h$  but, under the prior, are made using information through time  $t$ . The prior is incomplete, in the sense that  $\mathcal{D}$  only forms an expectation of the product of the agent densities. Agent opinion analysis theory (West and Crosse, 1992; West, 1992), extended to a time-series context by McAlinn and West (2019), shows that the posterior conditional density for  $y_{t+h}$  under this incomplete prior takes the form

$$p(y_{t+h} | \Psi_{t+h}, \mathcal{H}_t) = \int \alpha(y_{t+h} | \mathbf{x}_{t+h|t}, \Psi_{t+h}) \prod_{j=1}^J \pi_{jt}(x_{jt+h|t}) dx_{jt+h|t}, \quad (3.1)$$

where  $\alpha(y_{t+h} | \mathbf{x}_{t+h|t}, \Psi_{t+h})$  denotes the synthesis function that reflects how  $\mathcal{D}$  combines her prior information with the set of expert-based forecasts;  $\Psi_{t+h}$  denotes a matrix of parameters and latent states that control the properties of the synthesis function,  $\alpha(\cdot)$ .

### BPS in Practice

Theory offers no guide as to the specific choice of the synthesis function,  $\alpha(y_{t+h} | \mathbf{x}_{t+h|t}, \Psi_{t+h})$ . But a common choice in empirical applications, used, for example, in McAlinn and West (2019), McAlinn *et al.* (2020), and Aastveit *et al.* (2023), is to assume a dynamic linear regression model treating the draws from the  $J$  competing densities as generated regressors,  $\mathbf{x}_{t+h|t}$ . Our synthesis functions will have a dynamic regression form, but we will use a non-centered parameterization (see Frühwirth-Schnatter and Wagner, 2010):

$$\alpha(y_{t+h} | \mathbf{x}_{t+h|t}, \Psi_{t+h}) = \mathcal{N} \left( y_{t+h} | c_{t+h} + \sum_{j=1}^J (\gamma_j + \beta_{jt+h}) x_{jt+h|t}, \sigma_{t+h}^2 \right), \quad (3.2)$$

where  $c_{t+h}$  is a time-varying intercept assumed to follow a random walk,  $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_J)'$  are time-invariant weights, and  $\boldsymbol{\beta}_{t+h} = (\beta_{1t+h}, \dots, \beta_{Jt+h})'$  denotes the time-varying combination weights. As discussed above, a common choice in the literature is to assume that the weights,  $\beta_{jt+h}$ , evolve as a random walk (RW) with innovation covariance matrix  $\mathbf{V}$ , leading to a version of BPS that we label “BPS-RW.” When implementing BPS-RW in our empirical applications below, we make standard choices for the prior

and MCMC method. In particular, they are similar to those used in Hauzenberger *et al.* (2022). The only difference is that we use the hyperparameter-free horseshoe prior instead of the normal-Gamma prior, so as to have a prior that is comparable to the one used with our regression-tree model. Accordingly, we do not provide additional details here on drawing the time-varying weights for BPS-RW; see Hauzenberger *et al.* (2022) for details.

In all of our implementations of BPS, including BPS-RW, we consider two versions: one that assumes stochastic volatility (SV) and another that is homoskedastic. In the SV case, the error variance,  $\sigma_{t+h}^2$ , changes over time. We assume that the log-volatilities  $\varsigma_{t+h} := \log \sigma_{t+h}^2$  evolve according to an AR(1) model with autoregressive coefficient  $\rho_\varsigma$ , unconditional mean  $\mu_\varsigma$ , initial value  $\varsigma_0$ , and error variance  $\sigma_\varsigma^2$ . The prior choices for these parameters are given in Appendix B.1.1. Homoskedastic cases are obtained setting  $\sigma_\varsigma^2$  to zero. Below, for notational ease, we do not explicitly note those parameters relating to SV in the conditioning arguments.

All of our implementations of BPS also include a time-varying intercept,  $c_{t+h}$ , which is assumed to follow a random walk. As discussed below,  $c_{t+h}$  is included to allow for model incompleteness. Further econometric details are provided in Appendix B.1.

With these notational conventions established,  $\Psi_{t+h} = \left( \gamma, \{c_\tau, \beta_\tau, \sigma_\tau\}_{\tau=0}^{t+h}, \theta \right)$ , where  $\theta$  will be method-specific parameters that define the law of motion of latent states or appear in the hierarchical priors (such as  $\mathbf{V}$  in the case of BPS-RW).

The synthesis function,  $\alpha(y_{t+h} | \mathbf{x}_{t+h|t}, \Psi_{t+h})$ , is quite flexible, given that the weights it attaches to each of the  $J$  densities are dynamic and because it allows for time-varying error variances. We can also see that while Eq. (3.2) implies a Gaussian density conditional on  $\beta_{t+h}$ ,  $\mathbf{x}_{t+h|t}$ , and  $\sigma_{t+h}^2$ , when carrying out predictive inference we marginalize out the unknowns of the model, leading to a predictive density that can be highly non-Gaussian; see Eq. (3.3) below.

In contrast with other approaches to combining models and density forecasts, such as Bayesian model averaging (BMA), the weights on each density are restricted neither to lie between zero and one nor to sum to unity. In the case of BPS-RW, the degree of change in the weights will depend on the magnitude of the state innovation variances

for these parameters: small values imply slow, smooth adjustment of the weights over time, while large values allow for bigger, sharper changes.

Two additional aspects of this parameterization of the synthesis function are worth noting before we introduce our regression-tree approach, which provides a more flexible nonparametric representation of the synthesis function.

First, as a special case, we define a static version of BPS that assumes time-invariant weights  $\beta_\tau = \mathbf{0}_J$  for all  $\tau$  but leaves  $\gamma$  unrestricted. We label this instance of BPS, which assumes the combination weights to be constant over time, “BPS-CONST.”

Second, the presence of both an intercept and an error in the synthesis function means that these versions of BPS allow for model set “incompleteness” (Geweke, 2010). That is, they allow the “true” (but unknown) model not to be in  $\mathcal{D}$ ’s model space; see, for example, Billio *et al.* (2013) and Aastveit *et al.* (2018). A conventional model combination scheme such as BMA sets both intercept and error to zero. The fact that the intercept,  $c_{t+h}$ , and error variance,  $\sigma_{t+h}^2$ , are both time varying provides additional flexibility when modeling the degree of model set incompleteness. Note that these specific assumptions are equivalent to embedding a popular benchmark for forecasting (especially of inflation) – the unobserved components SV (UCSV) model of Stock and Watson (2007) – within our set of now  $J + 1$  density forecasts. This is also related to an alternative treatment of model set incompleteness in BPS that adds a fictitious baseline predictive density to the set of densities being synthesized (see, for instance, the discussion in Section 2.2.3 of Tallman and West, 2023).<sup>4</sup> In our case, this baseline predictive density comes from a UCSV model. But importantly, as when estimating a mixture density, the parameters of the UCSV density are estimated simultaneously with the weights in the synthesis function.

To carry out predictive inference, we need to compute the predictive distribution. We do so by simulation. Let  $y_{T+h}$  denote a future realization of our target variable at time  $T + h$  and let  $\mathcal{H}_T$  denote the set of agent densities that are available at time  $T$

---

<sup>4</sup>Diebold *et al.* (2023) also add a fictitious forecaster in their ECB SPF application that, like ours below, combines forecaster-level density forecasts.

but target  $T + h$ . The predictive density, in our case, is obtained as follows:

$$p(y_{T+h}|\mathcal{I}_T) = \int \int p(y_{T+h}|\Psi_{T+h}, \mathcal{H}_T) d\Psi d\mathcal{H}, \quad (3.3)$$

where  $\mathcal{I}_T$  indicates the information set up to time  $T$  and  $\Psi_{T+h}$  are the latent states (projected forward to time  $T + h$ ). We can simulate from (3.3) by simulating from the joint posterior of the agents and states, projecting the states forward to time  $T + h$ , and then using the synthesis function in (3.2) to produce a combined forecast draw. By doing so, we integrate out the unknowns of the model, and the resulting predictive density can be highly non-Gaussian and feature heavy tails, multi-modalities, and/or skewness.

### 3.2.2 Bayesian Predictive Synthesis with Regression Trees (BPS-RT)

In this paper, our proposal is to relax the restrictions in BPS-RW by considering more flexible forms of time variation in  $\beta_{t+h}$ . Specifically, we use techniques from machine learning to model the dynamic evolution of the weights,  $\beta_{t+h}$ , in a nonparametric manner as a function of additional weight modifiers. This treatment can be contrasted with the alternative of treating the function,  $\alpha$ , itself nonparametrically. We follow Chipman *et al.* (2010) and use Bayesian additive regression trees (BART) to estimate the regression trees. BART consists of a set of priors for the tree structure and the terminal nodes (the leaf parameters) and a likelihood for data in the terminal nodes.

BPS-RT differs from existing implementations of BPS through both the hierarchical priors used on elements in  $\gamma$  and  $\beta_{t+h}$  and the incorporation of additional covariates into  $\mathcal{D}$ 's information set. These are stored in a  $K_\gamma$  vector  $\mathbf{z}_j^\gamma$  and a  $K_\beta$  vector  $\mathbf{z}_{jt+h|t}^\beta$ , both containing additional ‘‘data’’ known to  $\mathcal{D}$  through period  $t$ .

We postulate a nonlinear relationship between the weights and the weight modifiers through functions  $\mu_j^\gamma(\mathbf{z}_j^\gamma)$  and  $\mu_j^\beta(\mathbf{z}_{jt+h|t}^\beta)$  that determine the state transition equation that can be interpreted as a prior. In particular, we assume

$$\gamma_j \sim \mathcal{N}(\mu_j^\gamma(\mathbf{z}_j^\gamma), \tau_j^\gamma) \quad \text{and} \quad \beta_{jt+h} \sim \mathcal{N}(\mu_j^\beta(\mathbf{z}_{jt+h|t}^\beta), \tau_j^\beta), \quad (3.4)$$

where  $\tau_j^\gamma$  and  $\tau_j^\beta$  denote prior scaling parameters. For convenience, we define  $\mu_j^\gamma := \mu_j^\gamma(\mathbf{z}_j^\gamma)$  and  $\mu_{jt+h}^\beta := \mu_j^\beta(\mathbf{z}_{jt+h|t}^\beta)$ . The best way to illustrate the effect the scaling parameters have on the actual estimates of the weights is to consider a reparameterization of the synthesis function. Integrating out  $\gamma_j$  and  $\beta_{jt+h}$  by plugging Eq. (3.4) into Eq. (3.2) yields

$$y_{t+h} = c_{t+h} + \sum_{j=1}^J \left[ \underbrace{\left( \mu_j^\gamma(\mathbf{z}_j^\gamma) + \sqrt{\tau_j^\gamma} \nu_j^\gamma \right)}_{\gamma_j} x_{jt+h|t} + \underbrace{\left( \mu_j^\beta(\mathbf{z}_{jt+h|t}^\beta) + \sqrt{\tau_j^\beta} \nu_{jt+h}^\beta \right)}_{\beta_{jt+h}} x_{jt+h|t} \right] + \sigma_{t+h} u_{t+h}, \quad (3.5)$$

with  $\nu_j^\gamma, \nu_{jt+h}^\beta \sim \mathcal{N}(0, 1)$  denoting process innovations. The innovations,  $\nu_j^\gamma$  and  $\nu_{jt+h}^\beta$ , and the corresponding scaling terms control the degree of dispersion of the actual weights from those expected under the prior mean. If the scalings are close to zero, the posterior of  $\gamma_j$  and  $\beta_{jt+h}$  is pulled toward the prior mean and the resulting estimates will be close to  $\mu_j^\gamma(\mathbf{z}_j^\gamma)$  and  $\mu_j^\beta(\mathbf{z}_{jt+h|t}^\beta)$  and so strongly depend on  $\mathbf{z}_j^\gamma$  and  $\mathbf{z}_{jt+h|t}^\beta$ . If this is not the case, the resulting scaling parameters would be larger, so that substantial deviations from the prior means are more likely. Another feature of this representation is worth emphasizing. As opposed to a model that directly approximates the synthesis function nonparametrically, the specification in (3.5) introduces interaction terms of the form  $\mu_j^\gamma(\mathbf{z}_j^\gamma) \times x_{jt+h|t}$ . This specific form might reduce the risk of overfitting by introducing more structure on the space of functions that we approximate.

We approximate the prior mean functions through a sum-of-trees model with  $S$  trees:

$$\mu_j^\gamma(\mathbf{z}_j^\gamma) \approx \sum_{s=1}^S g(\mathbf{z}_j^\gamma | \mathcal{T}_s^\gamma, \phi_s^\gamma) \quad \text{and} \quad \mu_{jt+h}^\beta = \mu_j^\beta(\mathbf{z}_{jt+h|t}^\beta) \approx \sum_{s=1}^S g(\mathbf{z}_{jt+h|t}^\beta | \mathcal{T}_s^\beta, \phi_s^\beta), \quad (3.6)$$

where  $g$  denotes a tree function that is parameterized by so-called tree structures,  $\mathcal{T}_s^n$ , and terminal node parameters,  $\phi_s^n$ , for  $n \in \{\beta, \gamma\}$ . The basic idea behind a single tree is that the tree structures describe sequences of disjoint sets. These sets partition the input space (determined by exogenous covariates,  $\mathbf{z}_j^\gamma$  and  $\mathbf{z}_{jt+h|t}^\beta$ , respectively). Each of these sets is associated with a particular terminal node parameter. In our case, the

terminal node parameters serve as prior expectations for the  $\gamma_j$ s and for the  $\beta_{jt+h}$ s. The input space is associated with vectors of variables,  $\mathbf{z}_j^\gamma$  and  $\mathbf{z}_{jt+h|t}^\beta$ , which we refer to as weight modifiers. Note that  $\mathbf{z}_{jt+h|t}^\beta$  could include quantities (such as moments from the agent-specific predictive densities) that explicitly target  $t+h$  but are available at time  $t$ .

There are two main justifications for our BPS-RT modeling approach. First, there is no reason to restrict attention, as in BPS-RW, to random walk specifications for the evolution of  $\beta_{t+h}$ . BPS-RW implies, at a given point in time, a linear relationship between  $y_{t+h}$  and  $\mathbf{x}_{t+h|t}$ . This assumption might be warranted in tranquil periods. But, in unusual times, nonlinearities could be present, and exploiting these might lead to more accurate forecasts. Our regression-tree approach allows for flexibility in the way such nonlinearities are modeled and lets the “data speak.” Second, and this holds across all existing instances of BPS not just BPS-RW, an implicit assumption made is that the information set available to  $\mathcal{D}$  comprises exclusively the agent-based forecast densities.<sup>5</sup> But, in principle, additional unmodeled information is available to  $\mathcal{D}$  and might help inform evolution of the weights. In our BPS-RT approach, the weight modifiers,  $\mathbf{z}_j^\gamma$  and  $\mathbf{z}_{jt+h|t}^\beta$ , comprise this extra information.

These weight modifiers might include characteristics of the agents’ forecasts not directly reflected in their predictive distributions or other common (to agents) factors, such as general information about the macroeconomic environment. For example,  $\mathbf{z}_j^\gamma$  might contain summary metrics of overall past forecast performance (such as the average historical forecast performance) for each agent. Or, as noted above,  $\mathbf{z}_{jt+h|t}^\beta$  might contain more granular and time-varying information, such as time-varying characteristics of the agent-specific predictive densities (say their higher moments and/or time-varying measures of past forecasting performance). We provide specific context and motivate our choice of weight modifiers in the empirical applications in Section 3.3.1 below.

To return to the regression tree, note that it is defined by disjoint sets that are determined by splitting rules of the form  $z_{k,jt+h|t}^\beta \leq d_k$  or  $z_{k,jt+h|t}^\beta > d_k$ , where  $z_{k,jt+h|t}^\beta$

---

<sup>5</sup>As mentioned in footnote 2, an exception is Oelrich *et al.* (2023), who, when combining density forecasts using the linear opinion pool, also let the weights depend on exogenous variables. Our BPS-RT model generalizes to consider BPS combinations beyond the linear special case and to allow for nonlinearities in how the weight modifiers affect the weights.

is the  $k^{\text{th}}$  weight modifier for the  $j^{\text{th}}$  agent/model and  $d_k$  is a threshold parameter associated with the  $k^{\text{th}}$  effect modifier, which is estimated from the data. It is important to note, however, that any splitting rule associated with the  $k^{\text{th}}$  effect modifier is common across agents and periods (that is, it is specific neither to agent  $j$  nor to period  $t$ ). Hence, the thresholds  $d_k$  and thus the tree structures do not have  $t$  or  $j$  subscripts and are common across agents/models and time. Since these splitting rules effectively govern the prior mean,  $\mu_{jt+h}^\beta$ , this structure in a sense captures the notion of a pooling prior and reflects the situation that  $\mathcal{D}$  decides on the weights associated to each of the different agents based on using additional factors  $\mathbf{z}_j^\gamma$  and  $\mathbf{z}_{jt+h|t}^\beta$  according to a set of common decision/splitting rules. The same structure also holds for the  $\gamma_j$ s, with the difference that the splitting rules controlling  $\mu_j^\gamma$  pool exclusively over the cross-section and not over time (since the  $\gamma_j$ s are time invariant).

To see this pooling feature more clearly, consider a BPS-RT model that assumes  $\beta_{t+h} = \mathbf{0}_J$  and features only a time-variant part  $\gamma$ , for which the prior mean  $\boldsymbol{\mu}^\gamma = (\mu_1^\gamma, \dots, \mu_J^\gamma)'$  is defined by a single tree ( $S = 1$ ) and by a single effect modifier in  $z_j^\gamma$  (that is,  $z_j^\gamma$  is a scalar with  $K_\gamma = 1$ ). In this case, the prior on  $\gamma_j$  can be written as

$$\gamma_j = g(z_j^\gamma | \mathcal{T}_s^\gamma, \phi_s^\gamma) + \sqrt{\tau_j^\gamma} v_j, \quad v_j \sim \mathcal{N}(0, 1). \quad (3.7)$$

If we now compute the difference between  $\gamma_j$  and  $\gamma_m$  for distinct agents,  $j \neq m$ , and assume that  $z_j^\gamma$  and  $z_m^\gamma$  are similar, in the sense that both imply the same decomposition of the input space and are thus located in the same terminal node of the tree, we end up with

$$(\gamma_j - \gamma_m) \sim \mathcal{N}(0, \tau_j^\gamma + \tau_m^\gamma). \quad (3.8)$$

This equation implies that if the tree suggests that the characteristics between agents are so similar that they are grouped together in the same terminal node, the same prior mean applies and the difference between prior means will be zero. The presence of the prior scaling parameters  $\tau_j^\gamma$  and  $\tau_m^\gamma$  will then allow for data-based testing of whether that restriction should be strongly enforced or not. Since both prior means would coincide, setting both  $\tau_j^\gamma$  and  $\tau_m^\gamma$  to values close to zero would induce a clustering of  $\gamma_j$  and  $\gamma_m$

around  $g(z_j^\gamma | \mathcal{T}_s^\gamma, \phi_s^\gamma) = g(z_m^\gamma | \mathcal{T}_s^\gamma, \phi_s^\gamma)$ . Hence, the choice of the prior specified on the scaling parameter  $\tau_j^\gamma$  is crucial in determining the clustering behavior of BPS-RT.

Another feature of our prior is that  $\mathcal{D}$  adjusts her weights on the agents' densities depending on the (common) macroeconomic environment as captured by the weight modifiers, which might include, as discussed, indicators of the state of the business cycle, measures of economic uncertainty, or deterministic trends. For example, in turbulent times, larger weights on component densities that are far from Gaussian and feature, say, heavy tails might lead to better combined density forecasts. Our approach can control for this, if supported by the data.

Note that we estimate the tree structures and the terminal parameters alongside all other unknown parameters and therefore also specify priors for them. We follow here the recommendations of Chipman *et al.* (2010) and discuss the remaining model and prior specification issues in detail in Appendix B.1. This technical appendix also describes the MCMC methods used to estimate BPS-RT. In summary, these MCMC methods are straightforward. They require a method for predictive simulation from each individual model (to draw from each agent's forecast density) and a method for drawing from the regression-tree model conditional on the individual-agent draws. For BPS-RT, the algorithm is taken directly from Chipman *et al.* (2010).

### 3.2.3 Illustrating BPS-RT

We now explain how BPS-RT works and allocates combination weights using an illustrative toy example. Assume that, unknown to  $\mathcal{D}$ , the "true" data for  $y_t$  are generated by the following threshold model:

$$y_t = \begin{cases} \rho_1 y_{t-1} + c\rho_2 y_{t-2} + \sigma_0 \nu_t, & \text{for } t = 1, \dots, 200 \\ c\rho_1 y_{t-1} + \rho_2 y_{t-2} + \sigma_0 \nu_t, & \text{for } t = 201, \dots, 350, \end{cases} \quad (3.9)$$

where  $\rho_1 = 0.8$ ,  $\rho_2 = -0.8$  and  $\sigma_0 = 1.2$ ,  $y_0 = y_1 = 0$ ,  $c = 1/100$ , and  $\nu_t \sim \mathcal{N}(0, 1)$ .

Then,  $J = 2$  agents predict  $y_t$  as follows (these forecasts are one-step-ahead,  $h = 1$ ):

$$x_{1t} \sim y_{1t} = \mathcal{N}(\rho_1 y_{t-1}, (1 - \rho_1^2) \sigma_0^2), \quad (3.10)$$

$$x_{2t} \sim y_{2t} = \mathcal{N}(\rho_2 y_{t-2}, (1 - \rho_2^2) \sigma_0^2). \quad (3.11)$$

Both agents use forecasting methods with a different type of misspecification. The first agent's forecast is almost correctly specified for the first part of the sample, but the second agent's is substantially misspecified. In the second part of the sample this switches. We would hope that BPS-RT, when combining these two misspecified densities, would put more weight on the first agent when  $t \leq 200$ , then increase the weight on agent 2 when  $t > 200$ .

Notice that the structure of the data-generating process (DGP) implies that BPS-RW is severely misspecified, since BPS-RW implies that the combination weights on the two agents evolve smoothly over time. Our more flexible choice of synthesis function, (3.2), conditional on choosing appropriate effect modifiers, as we shall show, is capable of accommodating the break at  $t = 200$ .

We consider three variables as weight modifiers. The first is a simple deterministic time trend,  $z_{1,jt+1|t}^\beta = t + 1$ , that is common to both agents. The remaining two effect modifiers are agent-specific and measure historical forecasting performance. To capture historical point forecasting performance, we consider each agent's squared forecast error (SFE) as recursively computed at time  $t - 1$ :  $z_{2,jt+1|t}^\beta = (y_t - \mathbb{E}(x_{jt|t-1}))^2$  for  $j = 1, 2$ . Then, to measure past density forecasting performance, we consider each agent's continuous ranked probability score (CRPS).<sup>6</sup>

Our synthesis function is given by Eq. (3.2). To facilitate illustration of BPS-RT, we make some simplifying assumptions. We set the time-invariant weights  $\gamma = \mathbf{0}$  and, for the prior on  $\beta_{t+1}$ , set the scaling parameters equal to zero so that the weights and prior means coincide, and we focus on the single-tree case ( $S = 1$ ). For expositional ease, we drop corresponding sub- and super-scripts when there is no loss in meaning. Under these simplifying assumptions, the synthesis function, similarly to (3.5), reduces

---

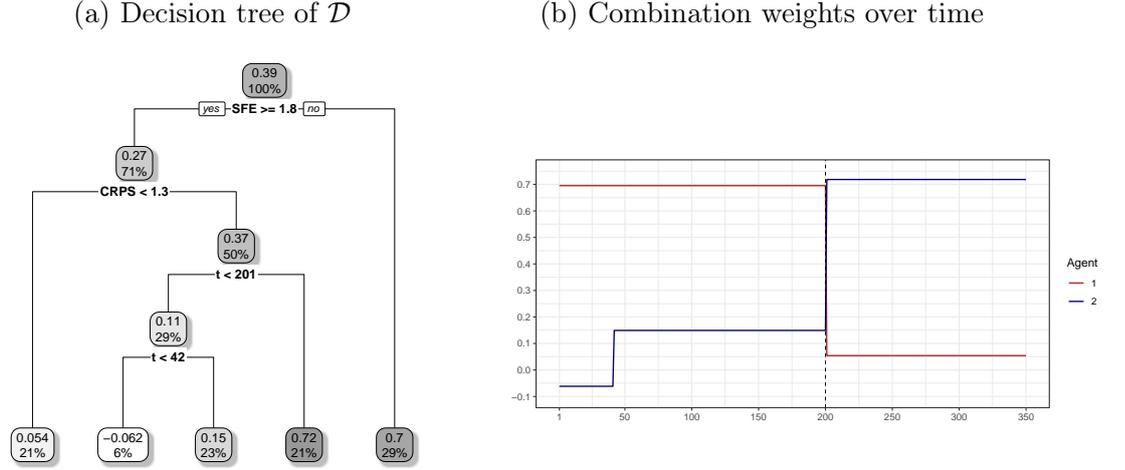
<sup>6</sup>If  $F$  is the c.d.f. of the forecast and  $y$  the subsequent realization, then  $\text{CRPS}(F, y) = \int (F(x) - \mathbf{1}_{x \geq y})^2 dx$ .

to

$$\alpha(y_{t+1} | \mathbf{x}_{t+1|t}, \mathbf{z}_{t+1|t}, \Psi_{t+1}) = \mathcal{N}(y_{t+1} | c_{t+1} + g(\mathbf{z}_1 | \mathcal{T}, \phi)x_{1t+1|t} + g(\mathbf{z}_2 | \mathcal{T}, \phi)x_{2t+1|t}, \sigma_{t+1}^2).$$

This equation shows that with the scaling parameters set equal to zero, we end up with a BART model that assumes the weights depend nonlinearly on  $\mathbf{z}_{t+1|t}$ .

Figure 3.1: Illustration of BPS-RT.



**Notes:** As the weight modifiers, we use a simple linear time trend, SFE, and CRPS. Each oval box in panel (a) indicates the terminal node parameter of a particular branch and the share (in percent) of observations belonging to this branch.

Figure 3.1 depicts in panel (a) the estimated tree and in panel (b) the temporal evolution of the estimated weights. We emphasize that these weights are in-sample estimates, that is, conditional on data through  $T = 350$ .

The tree in panel (a) can be understood as follows. Let us start at the bottom of the tree. We see five terminal nodes. Hence, we observe five groups/clusters that define the prior mean both over time and across agents. Put differently, there are five “breaks” over time and across agents in the prior mean.

How we pool is defined by the splitting rules. These are understood by turning to the top of the tree. At the root (level 0), the SFE is used as a splitting variable. The threshold parameter is 1.8 and, hence, if the SFE in  $t - 1$  is larger than or equal to 1.8, we move down the left branch of the tree. At the first level, the lagged CRPS shows up as the next threshold variable. If the CRPS is smaller than 1.3, we end up in a terminal

node and set the weight associated with an agent that has an SFE greater than or equal to 1.8 and a CRPS smaller than 1.3 equal to  $\mathbb{E}(\beta_{jt}) = 0.054$ . These conditions are fulfilled 21 percent of the time. By contrast, if the CRPS is greater than or equal to 1.3, we drop down to the second level of the tree. In this segment, time shows up as a splitting variable, and if  $t \geq 201$ , we assign a weight of 0.72. For  $t < 201$ , we introduce a further splitting rule that splits the sample once more by testing whether  $t < 42$ . If this is the case, a negative weight of  $-0.062$  is applied, whereas if  $42 \leq t < 201$  the weight is 0.15. If the past SFE is smaller than 1.8, we end up in the right branch of the tree and assign a weight equal to 0.7.

Hence, the tree suggests that, first and foremost,  $\mathcal{D}$  selects agents according to the past performance of their forecasts, since both SFE and CRPS are identified in the estimated tree. Under our DGP, this implies that weights dynamically update if a given agent issued a poor prediction in the previous period without taking into account the past performance of her forecasts. To understand how these decision rules translate into the actual evolution of model weights, panel (b) shows the weights over time. These indicate that in the first part of the sample, Agent 1 receives substantial weight, while Agent 2 receives relatively little weight. This makes sense, given that the former is only mildly misspecified, whereas the latter features substantial model misspecification. As expected, given the structural break in the DGP,  $\mathcal{D}$  now overweights the second agent, whereas the weight on Agent 1 is now much smaller.

This simple exercise illustrates how  $\mathcal{D}$  incorporates additional information (time and past forecast errors in this case) to combine models. In general, though, the prior scaling parameters in BPS-RT are greater than zero, and hence, the decision tree gives rise to prior expectations that in turn inform the posterior estimates of the weights. Hence, if there is no relationship between the weights and the weight modifiers, the resulting prior variance would be large and the weights would follow a white noise process.

### 3.3 Two Macroeconomic Forecasting Applications

We investigate the performance of BPS-RT in two forecasting exercises. In the first application, we combine predictive densities of GDP growth for the euro area (EA)

produced by individual professional forecasters participating in the ECB Survey of Professional Forecasters (SPF). Beyond its intrinsic interest, this data set is a good testing ground for BPS-RT because it has been used before when comparing alternative density forecast combination methods; see Diebold *et al.* (2023), Conflitti *et al.* (2015), and Chernis (2023). Second, we forecast US inflation using a set of autoregressive distributed lag (ADL) regression models. This data set and model set has been used by Stock and Watson (2003) and Rossi and Sekhposyan (2014), the latter using a similar ADL strategy to create each of the agent’s forecast densities.

These two applications differ not only geographically and in terms of target variables, but also in the number of agents and the nature of the forecast densities the agents provide. The EA GDP growth application features a relatively small number of subjective, most likely judgment-informed, forecasts (ECB, 2019) that are provided in the form of histograms (with  $J = 14$ ). In contrast, the US inflation application uses a large number of model-based predictive densities, which are continuous and produced with distinct ADL regressions (with  $J = 56$ ). Further details on the design of both applications are provided in the subsequent sub-sections 3.3.2 and 3.3.3. Both applications’ evaluation samples cover the global financial crisis, the euro area crisis, and the COVID-19 pandemic. Taken together, these two applications enable a comprehensive assessment of BPS-RT.

#### 3.3.1 BPS-RT Specifications

We experiment with several different specifications of BPS-RT to draw out how density forecast accuracy varies with the characteristics of the specific synthesis function used. In broad strokes, we look at the importance of time variation, in both weights and volatility, the number of trees, and the choice of weight modifiers. Accommodating temporal instabilities (for example, see Rossi, 2021) is important in macro-modeling, and so is a natural subject of inquiry, while the number of trees is an important aspect of specifying BART models. Being able to specify weight modifiers is an attractive feature of BPS-RT and allows the combination weights to change based on information exogenous to the individual agents but known to the BPS decision maker. Hence, this

is also a key area of inquiry.

We accordingly investigate the following four specifications of BPS-RT distinguished by their choice of weight modifier(s) and whether that choice introduces cross-sectional (which we label C) or cross-sectional and time variation (which we label TC) in the combination weights seen in (3.2).

- **BPS-RT(C): AVG.-SCORES:** This specification uses as weight modifiers for the cross-sectional varying coefficients ( $\mathbf{z}_i^?$ ) measures of each agent’s historical (ex post) forecast accuracy. Specifically, to capture past point and density forecast accuracy, it considers model-specific averages of the mean squared forecast errors (MSEs) and the CRPSs, respectively. These averages are computed recursively to reflect information known only to  $\mathcal{D}$  in real time and could help the synthesis function distinguish between “good” and “bad” forecasters. This specification assumes constant weights over a given estimation window, although the weights are updated recursively through the evaluation period.
- **BPS-RT(TC): EXO.-IND:** This specification selects as weight modifiers application-specific exogenous indicators. These vary over time but not over the cross-section, implying that they are the same for each agent. These indicators are intended to provide a signal on the state of the economy, prompting BPS-RT to reweight the individual agents while simultaneously fostering a certain degree of synchronization among them. For example, during periods of high economic uncertainty, financial stress, or elevated inflation expectations, BPS-RT may weight a subset of models more heavily. In the EA SPF application, we use the European economic policy uncertainty (EPU) index of Baker *et al.* (2016), available via <https://www.policyuncertainty.com>. In recessions, their uncertainty measure rises; so, allowing the combination weights to depend on uncertainty enables them to move with the business cycle. In the US inflation application, we consider measures of inflation expectations and financial conditions. Both variables have been considered in the inflation-at-risk literature (López-Salido and Loria, 2020). Specifically, we consider households’ one-year-ahead

inflation expectations from the University of Michigan survey and, as a broad measure of financial conditions, the Chicago Fed’s national financial conditions index (NFCI). Both measures are available from the Federal Reserve Bank of St. Louis (<https://fred.stlouisfed.org>). In our empirical application, where we use a direct forecast design, we lag these exogenous indicators by the forecast horizon  $h$  to acknowledge the reality that we do not observe values for them in a future period,  $t + h$ , but only have information up to  $t$ . To catch any other time effects, in both applications we also consider a time trend ( $t = 1, \dots, T$ ).

- **BPS-RT(TC): FEATURES:** In addition to the scores discussed above, this specification considers statistical “features” of each agent’s predictive density, known to  $\mathcal{D}$  in real time. Specifically, we consider the first four moments of each agent’s predictive density and the cross-sectional dispersion of the agents. The latter is measured by the standard deviation (at time  $t$ ) across the  $J$  agents’ (models’) mean forecasts. Consideration of these features allows the density combination weights, in effect, to cluster to reflect the marginal properties of the individual forecasts and their disagreement. For example, it may well be that high (ex ante) uncertainty forecasters should be weighted similarly. Besides these features, we consider historical point and density forecast performance, measured by lagged MSEs and lagged CRPSs. For this RT(TC) specification, it is noteworthy that both score measures are used in such a way that the averages of these lagged scores impact the time-invariant part of the weights  $\gamma$ , while the plain lagged scores impact the weights through the time-varying part  $\beta_t$ .
- **BPS-RT(TC): ALL:** This specification includes all the previously discussed weight modifiers. By looking at the weight modifiers individually, and adding features sequentially, we can assess the marginal benefit of each weight modifier.

For each of these four versions of the model, we consider models with SV and homoskedastic errors and we allow the BART specification to either have a single tree ( $S = 1$ ), leading to a Bayesian regression-tree specification (see Chipman *et al.*, 1998), or a large number of trees ( $S = 250$ ), leading to BART. In traditional Bayesian im-

plementations using trees for nonlinear regression, such as Chipman *et al.* (2010), it is generally found that increasing the number of trees, starting at  $S = 1$ , leads to an improvement in forecast performance. But this improvement tends to peter out when the number of trees gets moderately large. The conventional wisdom is that the precise choice of the number of trees is not that important, provided that too small a value is not chosen. This may not be the case in BPS, since the data may prefer to have weights that are reasonably constant over time and change only occasionally. Hence, we choose to focus on single-tree specifications and BART to model the weights in BPS. As we shall see, we find that single-tree methods tend to forecast more accurately. As benchmarks in the forecasting exercises below, we consider both BPS-CONST and BPS-RW (as defined in Section 2.1.2).

### 3.3.2 Forecasting EA Output Growth Using the Survey of Professional Forecasters

The ECB has been producing the SPF since 1999. The ECB SPF is the longest-running EA survey of macroeconomic forecasts. Each quarter, the survey elicits from a panel of professional forecasters point and probability forecasts of EA inflation and GDP growth at various horizons.<sup>7</sup> We consider the two-quarter-ahead forecasts of year-on-year EA GDP growth. On average, there are 50 responses a quarter from a survey panel of over 100 professional forecasters.

There are a couple of features of the forecaster-level density forecasts from the ECB SPF that we have to address in order to combine them. First, survey respondents provide their probability forecasts over given (fixed) ranges. That is, they produce histogram rather than continuous density forecasts. For example, in the 1999Q1 survey, forecasters were instructed to provide their probability forecasts over 10 bins. The first bin was GDP growth less than 0 percent, with the bins then increasing in intervals of 50 basis points, until the tenth bin of higher than 4 percent growth. To accommodate the discretized nature of these probability forecasts, rather than fit a continuous density to the histogram (that may or may not have a good fit), we use the histogram forecast data

---

<sup>7</sup>For a full description of the EA SPF, see García (2003).

“as is.” We do this by, within our BPS approach, drawing samples for each forecaster directly from the histograms. Details of our algorithm, which involves a Metropolis-Hastings step, are given in Appendix A.2. Our sampling approach changes over time to capture the fact that the bin definitions have been moved over time. In particular, after shocks such as the global financial crisis and COVID-19, the ECB shifted the bins to allow forecasters to say more about the probabilities in what were, prior to the survey change, the extremes of the distribution. We also have to take a stand on the open intervals at the bottom and top of the histogram. We set the end-points for the histograms equal to the outer bin plus or minus (depending on whether we are at the top or bottom of the histogram) two standard deviations of GDP growth, as estimated using the vintage of GDP data available at the time the forecast was made.

Second, forecasters enter and exit the panel. This means that the panel is unbalanced. We follow Diebold *et al.* (2023) in constructing the longest consistent panel possible by dropping forecasters who are regular non-responders and then filling in the occasional missing values for the remaining forecasters. Specifically, we drop forecasters who have not responded for five or more consecutive quarters. This results in a panel of 14 forecasters. Any missing forecast data for these 14 forecasters are estimated using a Normal distribution based on the unconditional distribution of GDP growth as estimated in real time.<sup>8</sup>

We then take these 14 forecasters’ densities and carry out a recursive out-of-sample evaluation of the alternative BPS specifications over the sample 2005Q2 through 2021Q1. To do this, we first estimate the BPS combinations on a set of training samples that comprise a sequence of expanding windows of GDP and density forecast data. The GDP data used in the training sample are that vintage of GDP data available to the forecasters when they made their forecasts. The first training sample uses forecasts from the five-year period targeting GDP outturns from 1999Q3 through 2004Q2. These forecasts are taken from the surveys administered between 1999Q1 and 2003Q4. Given its publication lags and our desire to approximate the information set available at the time the

---

<sup>8</sup>We differ from Diebold *et al.* (2023) in two ways. First, they interpolate missing forecasts based on historical performance. Second, we have a different number of forecasts because we use a different sample and we forecast GDP growth instead of inflation.

SPF forecasts are publicly available, the GDP outturns required to estimate the BPS synthesis function over this training sample are taken from the 2004Q4 vintage. This estimated synthesis function then uses the 2004Q4 survey to forecast (out-of-sample) 2005Q2. The training sample and vintage of GDP data are then extended by one quarter, and forecasts are produced for 2005Q3. This process is continued until forecasts are produced for 2021Q1. This set of out-of-sample BPS density forecasts is then evaluated against GDP outturns taken from the June 9, 2021, vintage.

### 3.3.3 Forecasting US inflation Using a Set of Indicators from FRED-QD

We follow Rossi and Sekhposyan (2014) and construct density forecasts of US inflation using a set of autoregressive distributed lag (ADL) models. Each ADL model considers 1 of 27 indicators taken from the FRED-QD data set (McCracken and Ng, 2021), which is commonly used when forecasting macroeconomic aggregates such as inflation in the US. The selected indicators capture movements in assets prices, measures of real economic activity, wages and prices, and money. This rich and diverse set of economic indicators allows the ADL density forecasts of US inflation to display significant heterogeneity. Table B.1 in the Appendix provides an overview of the variables used as exogenous predictors and the transformations applied to ensure their stationarity.

We then use each of these ADL models to produce direct forecasts for quarter-on-quarter consumer price (CPIAUCSL) inflation one quarter ahead ( $h = 1$ ) and one year ahead ( $h = 4$ ). Specifically, for each indicator,  $x_{jt}$ , for  $j = 1, \dots, 27$ , we estimate the set of ADL models:

$$\pi_{t+h} = \rho_\pi \pi_t + \alpha_\pi x_{jt} + \varepsilon_{\pi,t+h}, \quad \varepsilon_{\pi,t+h} \sim \mathcal{N}(0, \sigma_{\pi,t+h}^2), \quad (3.12)$$

where  $\pi_t$  is inflation,  $\rho_\pi$  is the autoregressive coefficient, and  $\alpha_\pi$  denotes the coefficient related to the  $j^{\text{th}}$  exogenous indicator.<sup>9</sup> We supplement these  $j = 1, \dots, 27$  models with a 28<sup>th</sup> model (the AR(1) model) that sets  $x_t = 0$  in Eq. (3.12). We also allow

---

<sup>9</sup>For notational ease, we do not use  $j$  subscripts to distinguish parameters in Eq. (3.12).

$\sigma_{\pi,t+h}^2$ , the error variance, to be both time-varying and constant. Hence, we estimate 28 models both with and without SV, delivering, in total, a set of 56 individual models whose density forecasts we then combine using BPS. All 56 models are estimated using standard Bayesian techniques. Details are provided in Appendix B.1.3.

We first estimate these models on a training sample from 1970Q1 to 1989Q4. We then iterate forward using a rolling estimation window of 80 quarters to account for possible structural changes in the US economy. The first ten years of forecasts (1990Q1 to 1999Q4) are used as a training window to estimate the BPS synthesis functions. The combined forecasts are then assessed on the evaluation sample 2000Q1 to 2022Q4. This evaluation period includes distinct economic periods characterized by different inflation dynamics, including the dotcom crash, the global financial crisis, the COVID-19 period, and the post-pandemic inflationary period.

### 3.3.4 Empirical Results

We break the empirical results into three parts presented in the following three subsections. First, we evaluate the relative and absolute density forecast accuracy of BPS-RT. Second, we examine why BPS-RT forecasts more accurately than the benchmarks by comparing features of their forecast densities. Third, we demonstrate aspects of interpretability of BPS-RT by examining how BPS-RT can be used to understand the role of model incompleteness, agent clustering, and the time-varying importance of the different effect modifiers.

#### Forecast Accuracy

We evaluate forecast accuracy in several ways. We first evaluate the point (conditional mean) forecasts, extracted from the combined densities, using the root mean squared forecast error (RMSE) loss function. Second, we evaluate the full predictive densities. We emphasize evaluation of the predictive densities rather than the point forecasts. Since the loss functions of forecast users tend not to be quadratic – as the density forecast literature (see Aastveit *et al.*, 2019) emphasizes – it is always important to produce and evaluate complete probabilistic forecasts. We measure the relative

forecast accuracy of the forecast densities using two popular metrics: CRPS and a tail-weighted CRPS. Both are loss functions that score the density forecast according to the realization that subsequently materializes. CRPS evaluates the “whole” density, while tail-weighted CRPS focuses on accuracy in the tails (Gneiting and Ranjan, 2011).<sup>10</sup> We also test the absolute calibration of the combined density forecasts using the Rossi and Sekhposyan (2019) test on the probability integral transforms (PITs); and we assess the temporal stability of forecast performance using the fluctuation test of Giacomini and Rossi (2010). The results of both these tests are summarized below, with full results presented in Appendix B.2.

Figures 3.2 and 3.3 report the relative forecast performance of the different models in the EA GDP growth and US inflation applications, respectively, using the RMSE, CRPS, and CRPS-tails loss functions. Each row in these figures reports the relative (to the BPS-RW benchmark) performance of the four BPS-RT specifications as differentiated by whether they use a single tree or 250 trees and whether they have SV or homoskedastic errors. The four columns in the figures refer to which set of weight modifiers is used.

Looking first at the RMSE panel in Figure 3.2 for EA GDP growth, we see little difference between the alternative BPS-RT specifications in terms of their point forecast accuracy. The accuracy of the BPS-RT specifications also tends to be similar to that of BPS-CONST and BPS-RW, with gains/losses in general only around 3 percent. This supports the stylized fact from the forecasting literature that equal-weighted combinations of point forecasts are hard to beat (see Timmermann, 2006b). Turning to US inflation (Figure 3.3), we do see in the RMSE panel that some of the tree-based methods now improve upon the point forecast accuracy of both benchmarks and in a manner that is statistically significant. Of particular note is the superior performance of the single-tree models, which almost always outperform the more complicated 250-tree models. We discuss this finding further below.

The CRPS panels in both Figures 3.2 and 3.3 reveal yet more of a payoff to using BPS-RT, certainly relative to BPS-RW, when we evaluate the whole density. Many of

---

<sup>10</sup>In the empirical appendix, we follow Gneiting and Ranjan (2011) and break CRPS tails into their left and right tails. See Figures B.1 and B.2 in Appendix B.2.

the forecast accuracy gains for BPS-RT are statistically significant. An implication of this finding is that BPS-RW's assumption that the combination weights follow a random walk is not supported by the data. But BPS-CONST, especially when BPS allows for SV, remains competitive for EA GDP growth.

The CRPS and CRPS tail results echo those under RMSE loss in concluding that single-tree structures,  $S = 1$ , are almost always preferred to  $S = 250$ . The fact that a single-tree model produces more accurate forecasts contrasts with the conventional wisdom in the wider BART literature; see Chipman *et al.* (2010). In our case, however, we model the weights, rather than the observed outcomes, nonparametrically; hence, the implied conditional mean relation (see Equation 3.5) introduces more restrictions relative to a standard BART model and hence lessens the risk of overfitting.

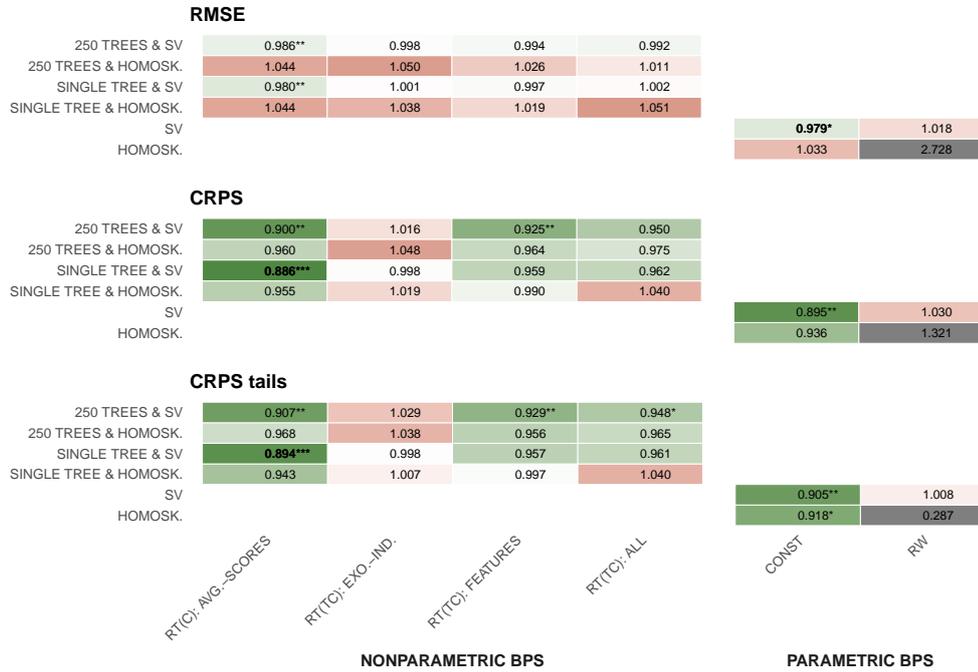
While the benefits of allowing for SV are well established in the density forecast literature (see Clark, 2011), allowing for SV in the BPS combination does not obviously improve the density forecasts from BPS-RT. But recall, and we touch on this again below when showing that these models in fact receive higher combination weights, in the US inflation application, half of the components models themselves allow for SV.

We now focus on comparing forecast accuracy across the first four columns of both Figures 3.2 and 3.3. This comparison reveals that the choice of weight modifier does affect forecast accuracy. It is not always the case that using more weight modifiers delivers more accurate forecasts. The benefit of different modifiers varies by application and by which row (which of the four BPS-RT specifications) is consulted.

Finally, we summarize the results from both the PITs calibration tests and the fluctuation tests. These results are reported in the online appendix for space reasons. The PITs plots (see Figure B.10) show that the BPS-RT densities are well calibrated, and especially so when forecasting EA GDP growth or US inflation one-quarter-ahead. The fluctuation test of Giacomini and Rossi (2010) reveals that there is temporal variation in the relative performance (under CRPS loss) of the preferred BPS-RT model and BPW-RW. Results (see Figure B.9) indicate that the superior performance of BPS-RT in the EA application is due to better forecasting performance toward the end of the global financial crisis. For US inflation, the better accuracy of BPS-RT is explained by

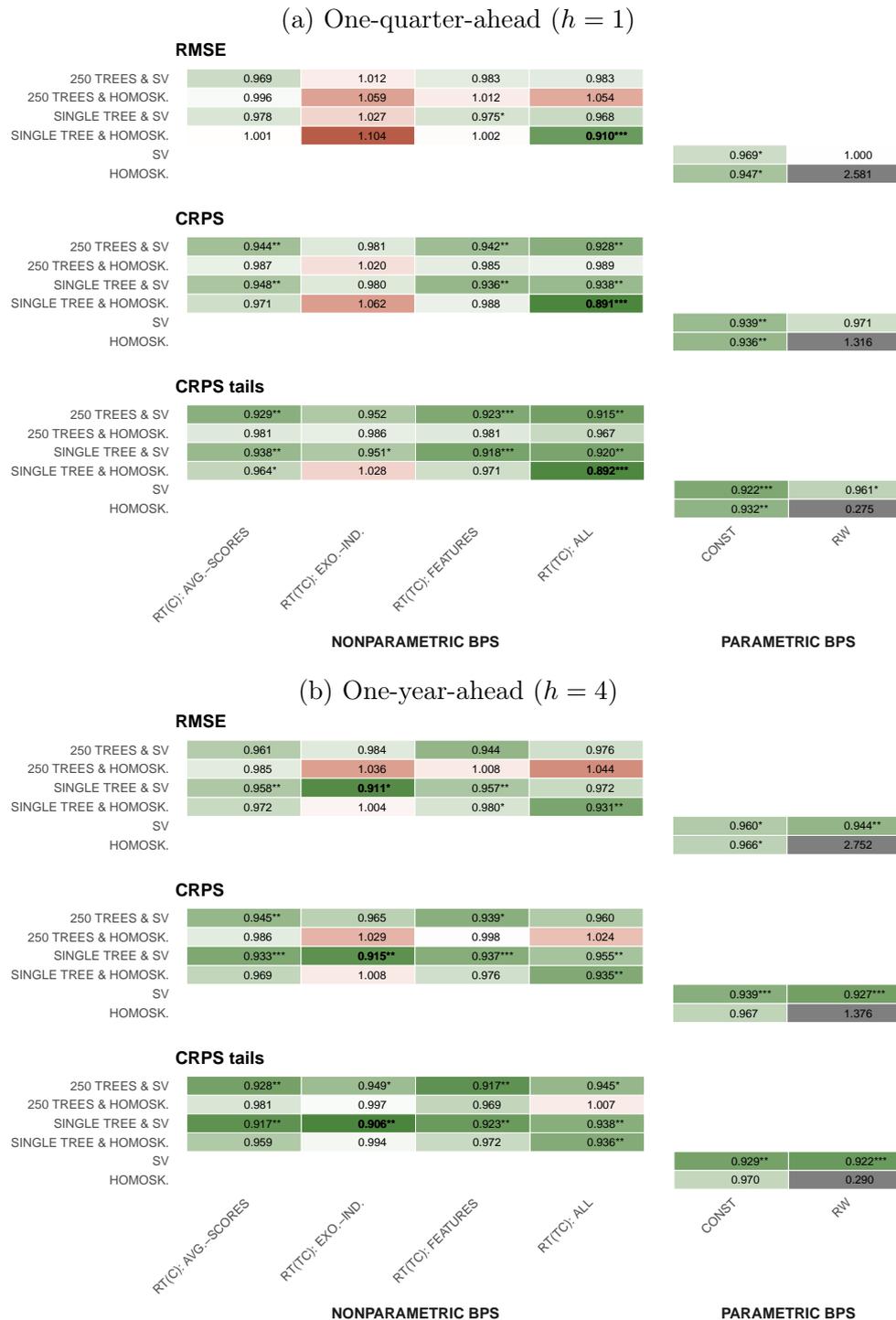
its more accurate density forecasts in the post-lockdown inflationary period.

Figure 3.2: Relative forecast accuracy: EA GDP growth.



**Notes:** This figure shows root mean square error (RMSE) ratios, (unweighted) continuous ranked probability score (CRPS) ratios, and a variant of quantile-weighted CRPS ratios that focuses on the tails. The gray-shaded entries give the actual scores of our benchmark (BPS-RW with homoskedastic error variances). Green-shaded entries refer to models that outperform the benchmark (with the forecast metric ratios below one), while red-shaded entries denote models that are outperformed by the benchmark (with the forecast metric ratios greater than one). The best-performing model specification by forecast metric is given in bold. Asterisks indicate statistical significance of the Diebold and Mariano (1995) test, which tests equal forecast performance for each model relative to the benchmark, at the 1 (\*\*\*) , 5 (\*\*), and 10 (\*) percent significance levels.

Figure 3.3: Relative forecast accuracy: US inflation.



**Notes:** This figure shows root mean square error (RMSE) ratios, (unweighted) continuous ranked probability score (CRPS) ratios, and a variant of quantile-weighted CRPS ratios that focuses on the tails. The gray-shaded entries give the actual scores of our benchmark (BPS-RW with homoskedastic error variances). Green-shaded entries refer to models that outperform the benchmark (with the forecast metric ratios below one), while red-shaded entries denote models that are outperformed by the benchmark (with the forecast metric ratios greater than one). The best-performing model specification by forecast metric is given in bold. Asterisks indicate statistical significance of the Diebold and Mariano (1995) test, which tests equal forecast performance for each model relative to the benchmark, at the 1 (\*\*\*) , 5 (\*\*), and 10 (\*) percent significance levels.

### Properties of the BPS-RT Density Forecasts

In this section we examine how and why BPS-RT forecasts more accurately. We focus on the best-performing (most accurate) model in each application and compare its forecast densities to those of the benchmark model, BPS-RW.<sup>11</sup>

Figure 3.4 shows a heat map of the difference in probabilities, in intervals of 1.5 percentage point for EA GDP growth and of 1 percentage point for US inflation, between BPS-RT and BPS-RW. Green (red) shading indicates that BPS-RT adds (subtracts) probability relative to BPS-RW in that interval. This is the approach pioneered by Diebold *et al.* (2023) as a way of visualizing the differences between competing density forecasts.<sup>12</sup>

Panel (a) of Figure 3.4 shows that, in general, BPS-RT predictions are less dispersed than BPS-RW with more mass near the subsequent outcomes. Additionally, the BPS-RT density adds probability to low GDP growth outturns prior to the financial crisis and also forecasts higher growth than BPS-RW in both the post-global financial crisis recovery and the rebound from the COVID-19-induced recession.<sup>13</sup> Panels (b) and (c) of Figure 3.4 show the analogous plots for US inflation. Similar to panel (a), BPS-RT places more mass closer to the outturn and produces forecasts that are, in general, less disperse. Moreover, BPS-RT adjusts much more quickly to the increase in inflation post-pandemic, both one quarter ahead and one year ahead, attributing a higher probability to these outturns than BPS-RW. Consistent with the evidence in Rossi and Sekhposyan (2014) that combinations of predictive densities for US inflation appear to be approximately Gaussian, the inflation forecast densities from BPS-RT also tend to

<sup>11</sup>As seen from Figures 3.2 and 3.3, in the EA GDP growth application, the “best” BPS-RT specification has a single tree and SV and uses average scores as effect modifiers (i.e.,  $RT(C): AVG. -SCORES$ ). For the US inflation application, the “best” BPS-RT specification has a single tree, homoskedastic errors, and the full set of weight modifiers (i.e.,  $RT(TC): ALL$ ).

<sup>12</sup>For an alternative but complementary visualization, Figure B.12 in Appendix B.2 shows the temporal evolution of the underlying density forecasts from BPS-RT and the benchmark BPS-RW model over the EA and US evaluation samples.

<sup>13</sup>As shown in Figure B.13 in the online appendix, in moving the probability mass from the centers to the left tail of the forecast density, BPS-RT captures asymmetries in the forecast densities. While there is some evidence of heightened downside risk asymmetries to GDP growth in the course of the financial crisis, consistent with the growth-at-risk literature (Adrian *et al.*, 2019), the evidence for negative skew is stronger still during the COVID-19 pandemic.

be symmetric (see Figure B.13 in the appendix), although there is clear evidence of a spike in downside risks in 2011, a time when the Fed was engaged in quantitative easing to combat deflation threats.

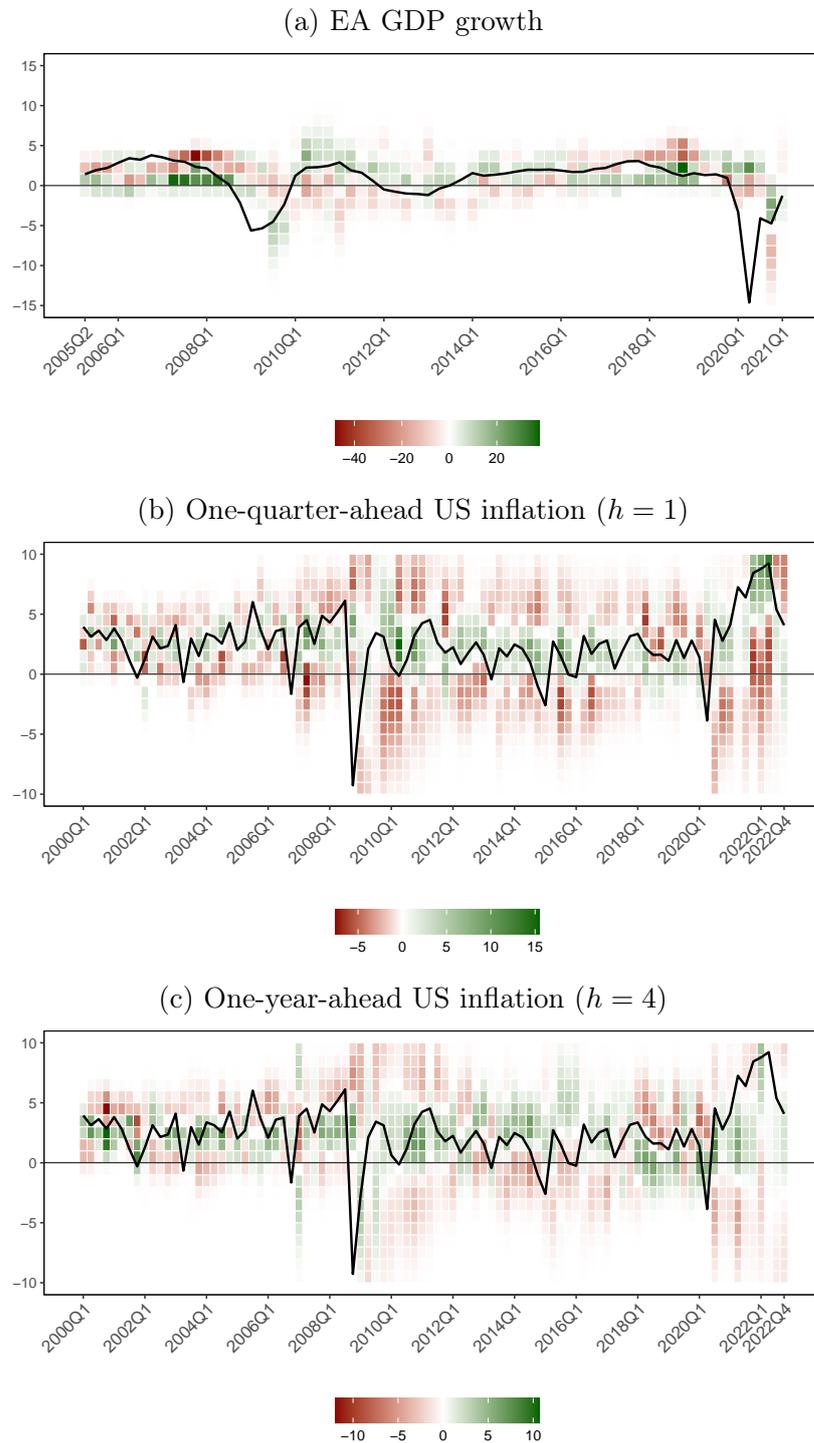
### **Interpretation: A Deeper Dive into the Mechanics of BPS-RT for US Inflation**

This section discusses how  $\mathcal{D}$  can interpret the combined forecasts from BPS-RT. In so doing, we continue to focus on the preferred BPS-RT specification in the US inflation application, not least because this is where we observe greater differences across the competing combination strategies. We first show how to quantify the degree of model set incompleteness, as a way of assessing how well the agents (the  $J$  forecasting models) that BPS-RT is combining are actually able to forecast. Second, we assess the relative importance of individual weight modifiers in driving BPS-RT.

To measure model set incompleteness, we compute an  $R^2$ -type measure. This estimates the proportion of the variation in  $y_{t+h}$  that is explained by the  $J$  agents. This measure is computed, for a specific period in the evaluation sample, as the ratio between the variation in the conditional mean in Eq. (3.2) explained exclusively by the BPS-RT component – which is the conditional mean in Eq. (3.2) without the time-varying intercept  $c_{t+h}$  – and the overall variation of the target variable,  $y_{t+h}$ .  $R^2$  values close to zero signify a high degree of model incompleteness, which means that the agents’ forecasts are not informative about the target variable. Instead, the intercept and error term in the BPS synthesis function, Eq. (3.2), explain a large portion of the total variation. In contrast,  $R^2$  values close to one indicate that the agents’ forecasts are informative and account for the majority of the variation, implying a complete model space.

Figure 3.5 plots this  $R^2$ -type estimate over the evaluation sample. Given that it is computed recursively, quarter by quarter, it experiences some volatility. But Figure 3.5 still evidences meaningful temporal variations in the degree of model set incompleteness at both forecast horizons. We see higher model incompleteness for the one-year-ahead forecasts than for the one-quarter-ahead forecasts. This is not surprising, as producing longer-horizon forecasts is obviously more difficult. At both horizons, we see increases in

Figure 3.4: Difference in probabilities between BPS-RT and BPS-RW

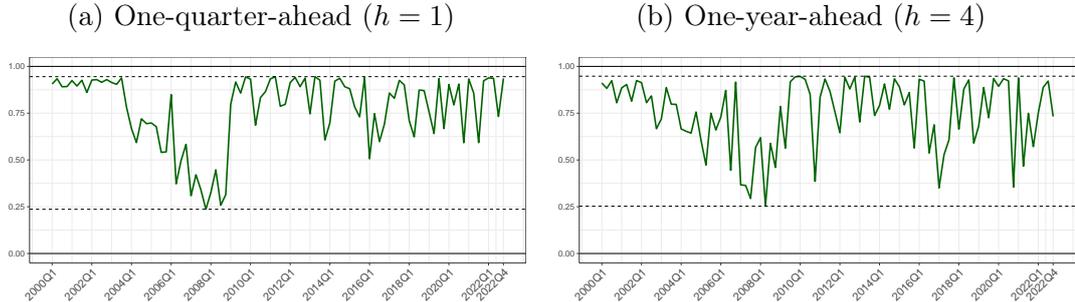


**Notes:** This figure shows the difference in probabilities between the best-performing BPS-RT model (in terms of CRPS) and BPS-RW. We define a grid of possible values for EA GDP growth ranging from  $-15$  percent to  $15$  percent with increments of  $1.5$  percent, while we define a grid of possible values for US inflation ranging from  $-10$  percent to  $10$  percent with increments of  $1$  percent. Green (red)-shaded cells indicate that the best-performing model adds (subtracts) probability relative to the benchmark in the respective region.

model incompleteness during the period 2004–2008, a time of extreme oil price volatility as well as the global financial crisis, and in the disinflation period after the 2015 oil price shock.

Interestingly, there is no clear evidence of an increase in model incompleteness during the post-pandemic rise in inflation, reinforcing the message from Figure 3.4 that BPS-RT was better able to anticipate the 2021 rise in US inflation.

Figure 3.5: Measuring model incompleteness: US inflation

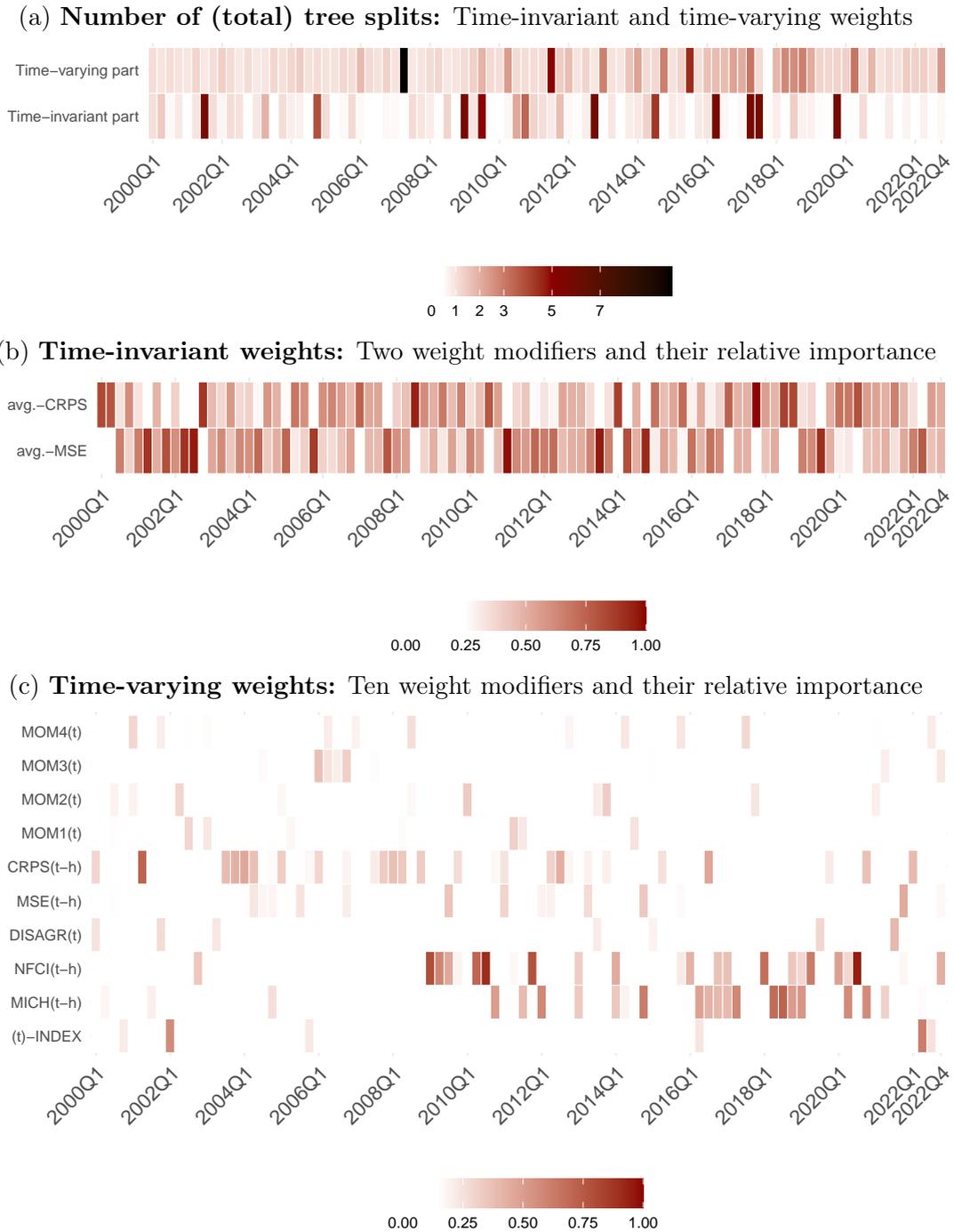


**Notes:** This figure shows the evolution of the model incompleteness measure over time. For each quarter in the evaluation sample, this measure is computed for our preferred specification (homoskedastic BPS-RT(TC): ALL with a single tree) as the ratio between the variation explained exclusively by the BPS-RT part (i.e., the conditional mean without the time-varying intercept) and the total variation, which thus can be interpreted as an  $R^2$  measure. The green solid lines represent the posterior median of this incompleteness  $R^2$ , which is bounded between zero and one. Values close to zero suggest that model incompleteness, as measured by the time-varying intercept and the error variance in Eq. (3.2), plays an important role, while values close to one indicate that the BPS-RT part explains most of the variation.

We now turn to assessing the relative importance of the individual weight modifiers in driving the density forecasts from BPS-RT. We do so by looking first at the number of tree splits and then by calculating inclusion probabilities for each weight modifier. Inclusion probabilities are calculated as the number of splits associated with the respective weight modifier divided by the total number of splits. For space reasons, we focus our discussion on Figure 3.6, which examines the weight modifiers for forecasting US inflation one quarter ahead. Analogous results forecasting inflation one year ahead are reported in Figure 3.7 and summarized below when the conclusions differ markedly from those discussed in greater detail for the one-quarter-ahead forecasts.

We start in panel (a) of Figure 3.6 by plotting the evolution of the total number of tree splits over the evaluation sample. This panel indicates whether variability in the combination weights comes from the time-varying ( $\beta_{jt+h}$ ) or constant component ( $\gamma_j$ ) of BPS-RT. Panel (a) reveals that BPS-RT tends to select a relatively small number of tree splits, especially for the time-invariant weights. Typically for  $\gamma_j$ , we observe that the posterior mean of the number of tree splits lies between 0.52 (lower quartile over the evaluation sample) and 1.15 (upper quartile, with a few more exceptions in the upper tail), while the average over the evaluation sample is 1.28. On the other hand, the posterior mean number of tree splits for the time-varying weights,  $\beta_{jt+h}$ ,

Figure 3.6: Number of tree splits for BPS-RT ( $S = 1$ ) and relative importance for each weight modifier for US inflation: One-quarter-ahead forecasts



**Notes:** Panel (a) shows the evolution of the total number of tree splits over time, while panels (b) and (c) show the marginal importance of each weight modifier for each quarter in the evaluation sample. Relative importance is defined as the share of the total number of splits. For each quarter in the evaluation sample, we obtain the posterior mean for these measures for our preferred specification (homoskedastic BPS-RT(TC): ALL with a single tree). For the exogenous indicators and the MSE/CRPS scores,  $(t - h)$  indicates that these measures are lagged by the forecast horizon  $h$ , while all other measures can be included contemporaneously.

ranges from 1.08 to 1.68 (indicating the interquartile range) and has an average of 1.59 over the evaluation sample. To place these numbers in the context of a single-tree split on, for example,  $\gamma_j$  indicates that the combination weights tend to cluster around two distinct prior means. With this in mind, we interpret the results in panel (a) as showing that the combination weights often fall into a handful of clusters that are more likely to be determined by time-specific factors. However, the number of splits is modest, so the weights are relatively stable over time. This finding is consistent with the density forecast combination literature that finds that constant weight combinations can forecast well (see, for example, Chernis, 2023).

Panels (b) and (c) of Figure 3.6 then show the inclusion probabilities for each of the constant and time-varying weight modifiers. Panel (b) shows the inclusion probabilities for the weight modifiers (CRPS and MSE) used to model the time-invariant combination weights. Neither CRPS nor MSE is obviously more important. Both weight modifiers receive positive and often fairly similar probabilities of inclusion. This implies that BPS-RT does partition models on the basis of their historical forecast accuracy.

Panel (c) of Figure 3.6 shows the importance of both time-varying weight modifiers. The first thing to notice is that there is much more sparsity in terms of the weight modifiers BPS-RT selects. In the first half of the evaluation sample, we see that features of the individual density forecasts drive the posterior inclusion probabilities. Specifically, we see that the moments of the marginal densities and CRPS, lagged by the forecast horizon  $h$ , are selected. But in the second half of the evaluation sample, we see the largest proportion of tree splits attributed to the NFCI during and immediately after recessions. The Michigan survey expectations measure also receives more weight after the financial crisis. This is evidence that nonlinear features of BPS-RT are driven by weight modifiers related to the business cycle. In other words, our BPS-RT model finds that the data support changing the combination weights abruptly with business cycle fluctuations. Finally, the time trend receives a higher weight in the post-COVID period of higher inflation seen in 2021 and 2022. This finding indicates that this inflationary episode – unprecedented within the sample – requires a substantial and rapid adjustment of the combination weights. These required weight dynamics cannot be fully captured

by the business cycle weight modifiers. Instead, a time trend (or, more precisely, a time dummy) is ideal for modeling such a regime shift from low to high inflation during this exceptional period.

Finally, we summarize the properties of the posterior median estimates of the combination weights that are plotted over the evaluation sample in Appendix B.2. We draw out two conclusions for the combination weights estimated when forecasting US inflation one quarter ahead (see Figure B.4). First, BPS-RT places more weight on those component models with SV, especially toward the end of the evaluation sample. This corresponds to the period when BPS-RT outperforms the BPS-RW benchmark (see Figure B.8).

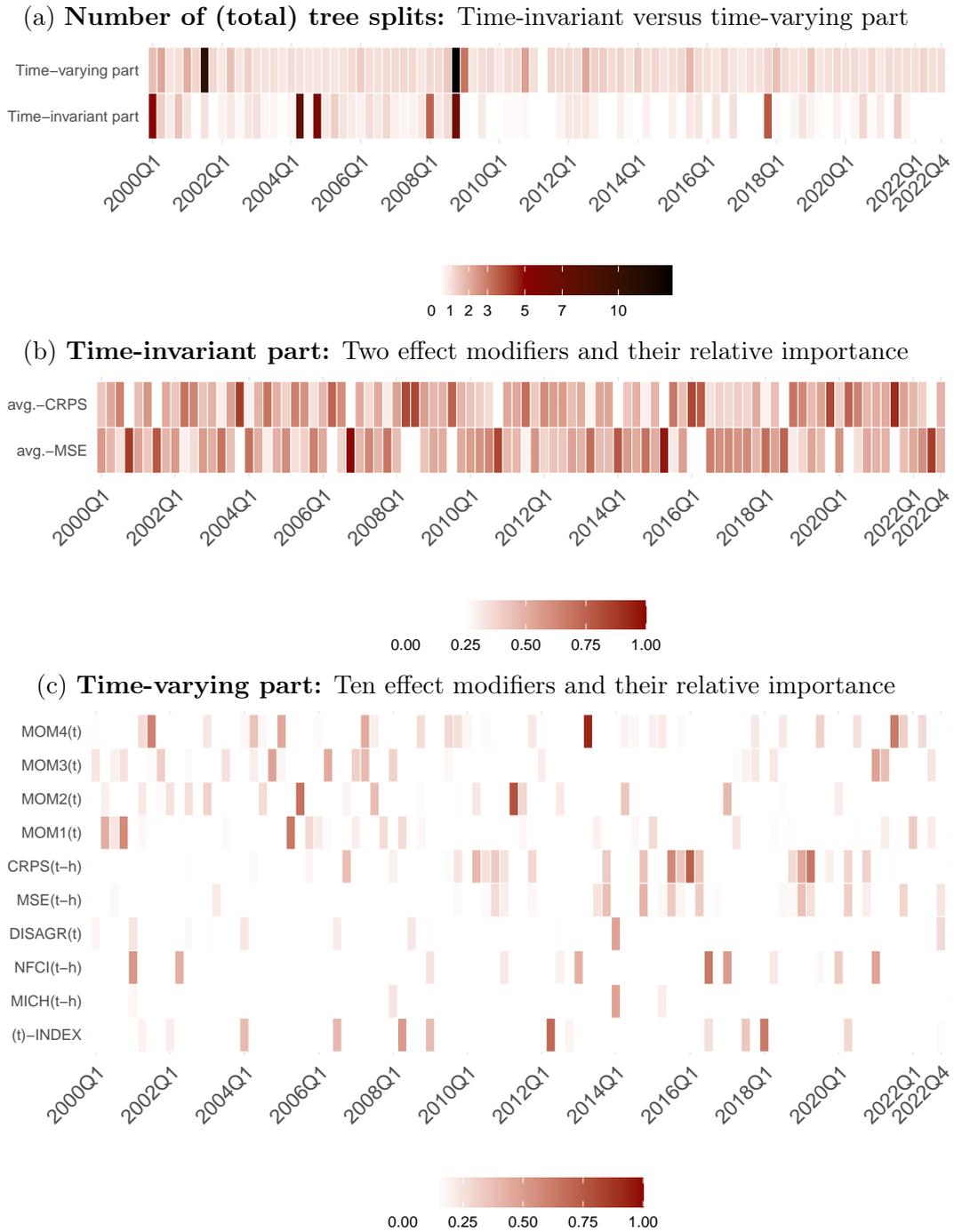
Second, among these SV models, only a subset receives large, in absolute value, weights. This indicates that there is some payoff, in terms of forecast accuracy, to occasionally placing a significantly higher weight on a small subset of models. Interestingly, some models get large negative weights. This amounts to short-selling those models as a “hedge” against the models with higher weights. A roughly similar pattern is seen for the one-year-ahead forecast combination weights seen in Figure B.5.<sup>14</sup>

While this subsection has focused on the US inflation application, we end by returning briefly to the EA GDP growth forecasting application. Figure B.3 in Appendix B.2, shows that the combination weights on most individual forecasters from the ECB SPF are, as anticipated given our earlier results, closer to equal than in the inflation application, where there was greater sparsity in the weights. This said, we do still see higher weights on a couple of experts (forecasters 6 and 14). We take this contrast-evidence across the two applications as empirical proof that BPS-RT is sufficiently flexible to adjust to forecasting scenarios that exhibit different dependence structures between the agents’ forecasts.

---

<sup>14</sup>Figure B.6 in the online appendix provides additional perspective on the temporal stability of the combination weights by plotting their sum over the evaluation sample. We see that when forecasting US inflation, this sum becomes negative during the global financial crisis, indicating how BPS-RT is reweighting most agents’ densities in the face of temporal instabilities. The sum of the weights also spikes upward during the 2021–22 inflationary episode, again indicating how BPS-RT can quickly adapt to temporal change.

Figure 3.7: **One-year-ahead horizon:** Number of (total) tree splits for our single-tree models and relative importance for each effect modifier.



**Notes:** Panel (a) shows the evolution of the total number of tree splits over time, while panels (b) and (c) show the marginal importance of each weight modifier for each period in the evaluation sample. Relative importance is defined as the share of the total number of splits. For each period in the evaluation sample, we obtain the posterior mean for these measures for our preferred specification (homoskedastic BPS-RT(TC): ALL with a single tree). For the exogenous indicators and the MSE/CRPS scores,  $(t-h)$  indicates that these measures are lagged by the forecast horizon  $h$ , while all other measures can be included contemporaneously.

### 3.4 Conclusion

Within the general BPS framework of McAlinn and West (2019), this paper develops a method for nonparametric density forecast combination using regression trees: BPS-RT. While a handful of papers use nonparametric techniques to combine densities, ours is the first to use regression trees. In contrast to most applications of regression trees, we model the coefficients, in our case the combination weights, instead of the variables using the regression trees. We show how this aids interpretation, since the combination model remains linear in the parameters. Additionally, regression trees use covariates, or weight modifiers, to drive changes in parameters, in contrast to conventional BPS applications where model parameters follow a random walk. Taken together, our approach is flexible but retains interpretability through linearity and the use of weight modifiers. We explain how BPS-RT can be used to understand the role of model incompleteness, agent (forecast) clustering, and the time-varying importance of the different weight modifiers.

We test the performance of BPS-RT in two different applications – combining model-based US inflation density forecasts and subjective histogram-based forecasts of euro area GDP growth. We find that, across both applications, BPS-RT forecasts well in terms of both relative and absolute accuracy. Interestingly, and in contrast to standard BART applications, we find that using a parsimonious single-tree specification outperforms models with more trees. Inspecting the best-performing specification, we observe that this superior performance is due to less disperse forecast densities and BPS-RT’s ability to better accommodate the shocks associated with the global financial crisis (in the GDP application) and COVID-19 (in the inflation application). Our proposed measure of model set incompleteness suggests that BPS-RT is able to capture much of the post-COVID rise in inflation. Triggered by a rise in the relative importance of the time trend in determining tree splits, itself highlighting the unusual nature of this inflationary period, BPS-RT also shifts its combination weights toward component models with SV. This contrasts with the prior period of lower inflation, when the business cycle indicators were found to be more important weight modifiers.

Future lines of research could involve investigating, in other forecasting applications and contexts, the usefulness of different sets of weight modifiers and the implications for weight structure. For instance, this could draw on the ability of BPS-RT, via its choice of weight modifiers, to capture general patterns of cross-sectional dependence between competing agents' probabilistic forecasts. Additional structure could be given to the clustering by, for example, letting the combination weight on a given individual agent's density forecast depend not only on characteristics of her own forecast (such as its mean or variance) but on characteristics of the other agents' forecasts.

## Chapter 4

# Decision Synthesis in Monetary Policy

### 4.1 Introduction

Monetary policymakers are tasked with simple, but hard to achieve, objectives. A common objective is to target the future inflation rate, or other macroeconomic outcomes, using interest rates as the policy instrument. These decisions are made based on uncertain information from many sources. In this paper, these sources are econometric models that generate predictive distributions for the macroeconomic outcomes and the policy instruments over multiple time periods. For a single model, it is straightforward to select an optimal policy instrument using decision analysis and conditional forecasting. Applying standard methods, such as Bayesian model averaging (BMA), is one way to address the issue of model uncertainty, as routine decision analysis can then be applied to the weighted average as a single model. However, this traditional view ignores the reality that a set of models may each individually recommend very different optimal policy decisions. The question then arises of how to synthesize this information and, potentially, exploit it in the overall final decision process. This paper addresses and answers this question.

There is an extensive Bayesian econometrics literature on model combination, but discussion of the issue that models are built for different purposes—specific prediction

and decision goals—is very sparse. Traditional BMA analysis weights models according to purely statistical model fit, and in time series explicitly, and only scores one step ahead of forecast outcomes. Extensions and alternatives have arisen to define model weightings based on aspects of past forecast performance with respect to specific forecast goals. Martin *et al.* (2023) survey Bayesian forecasting in economics and finance and review various forecast combination approaches, including some that are more explicitly concerned with goal-focused prediction (e.g., Mitchell and Hall, 2005; Geweke and Amisano, 2011; Conflitti *et al.*, 2015; Kapetanios *et al.*, 2015; Loaiza-Maya *et al.*, 2021; Chernis and Webley, 2022; Aastveit *et al.*, 2023; Bernaciak and Griffin, 2024). Lavine *et al.* (2021) provide additional perspectives and put many of the earlier approaches in a foundational Bayesian context. They justify model weights based on utilities in forecasting using historical model-specific “scoring” of past forecast outcomes. The underlying theoretical justifications come from Bayesian predictive synthesis (BPS) and the specific class of “mixture BPS” models (McAlinn and West, 2019, section 2.2; Johnson and West, 2022). However, while ultimate decision goals may be implicit in specific applications of model combination, they are rarely, if ever, taken into account in the analysis and resulting decision-making. This raises concerns. Although a model that has fit or forecast specific outcomes well in the past may be a good bet for use in resulting decision analysis—in our settings, defining optimal decisions about values of policy instruments—there is no guarantee that this will be so.

Our view is that models that have recommended policy decisions that turn out to be “good” should be more heavily weighted in looking ahead, just as past statistical predictive performance is generally positively weighted. The question is how to define “good,” since econometric models are not explicitly used and scored in past policy decisions. The challenge is then to operationalize the concept of “good decision” performance. For example, a vector autoregression (VAR) model can be evaluated on forecast performance using a pseudo real-time forecasting exercise. But, it is not clear how to evaluate such a model when it is used to advise policy decisions. We can, however, use the VAR model at present to inform near-term decisions and explore how it would have advised on decisions in the past. Evaluations can then compare such analysis to deci-

sions actually made by policymakers in the past (albeit recognizing that policymakers' past decisions were not necessarily correct and rather were just the outcomes of the amorphous reality of monetary policy-making).

Bayesian predictive decision synthesis (BPDS—Tallman and West, 2023; Tallman, 2024) addresses these questions. As part of the theoretical framework of Bayesian predictive synthesis (BPS—McAlinn and West, 2019; Johnson and West, 2022), BPDS explicitly allows and encourages the scoring of models based on decision analysis performance as well as statistical predictive accuracy. In addition to reflecting historical outcomes of predictions and decisions, BPDS critically also allows for differential model weighting based on *expected* decision outcomes. This is a complete decision parallel to the proven use of BPS models that incorporate outcome-dependent weights that modify BMA-like mixtures to differentially favour models in different parts of the future outcome space for pure forecasting. The latter concept was introduced by Kapetanios *et al.* (2015), whose empirically inspired developments recognized, for example, that one model may be better at predicting inflation when inflation is high and rising, while another model may be better when inflation is low and stable. BPS defines the conceptual and theoretical Bayesian bases and a broader methodological framework for this. BPDS goes further by integrating both historical and expected decision outcomes; here we develop, extend, and exemplify BPDS in our central macroeconomic policy context.

BPDS applies the broader Bayesian mixture model approach of BPS using defined utility—or “score”—functions that relate to explicit decision goals. Importantly, this allows for multiple objectives (i.e., multi-attribute decision analysis). For example, a purely predictive vector score function can allow for multiple forecast horizons (e.g., to produce inflation near a target for each of the next eight quarters) and/or multiple outcome criteria (e.g., to separately reflect inflation targeting, interest rate smoothing, and stable growth patterns over coming quarters), among others. For policymakers juggling multiple objectives, this key feature is rather distinct from conventional approaches that adopt single, scalar criteria for model weighting. For example, a forecast combination approach might choose model weights based on the  $h$ -step-ahead predictive likelihood for a single choice of  $h$ , with BMA simply focused on  $h = 1$ . In contrast, BPDS can

address multi-steps ahead in parallel, along with scoring the outcomes of decision goals that simultaneously target several macroeconomic outcomes.

The opportunities for exploring such practically relevant questions, and the differences relative to traditional single-model and BMA-based analyses, are showcased in our empirical studies. This involves an exploratory case study using US macroeconomic data with multiple-objective score functions to define BPDS model weights. This exemplifies the use of BPDS in macroeconomic forecasting and advisory decision-making.

## 4.2 BPDS Framework

We present and discuss the structure of BPDS at a particular point in time, ignoring the time dependency and relevance in the notation for clarity in communicating these essentials. Practical implementation in time series is of course sequential, with models at time  $t$  depending on all relevant historical data and information.

### 4.2.1 Mixture BPDS and Decision Setting

At a given time point, let  $\mathbf{y}$  denote the  $q$ -dimensional outcome variable of interest (e.g., inflation in each of the next  $q$  quarters) and  $\mathbf{x}$  the vector of control/decision variables (e.g., a target profile of central bank interest/base rates over the next  $q$  quarters). Each of a set of  $J$  models,  $\mathcal{M}_j$ ,  $j = 1:J$ , predicts the outcome  $\mathbf{y}$  via a predictive density  $p_j(\mathbf{y}|\mathbf{x}, \mathcal{M}_j)$  conditional on any considered decision  $\mathbf{x}$ . The policymaker responsible for ultimate decisions adopts a general BPDS approach with the overall conditional (on  $\mathbf{x}$ ) predictive pdf

$$f(\mathbf{y}|\mathbf{x}) \propto \sum_{j=0:J} \pi_j(\mathbf{x}) \alpha_j(\mathbf{y}|\mathbf{x}) p_j(\mathbf{y}|\mathbf{x}, \mathcal{M}_j) \quad (4.1)$$

with the following ingredients.

#### *BPDS model probabilities*

The *decision-dependent model probabilities*  $\pi_j(\mathbf{x})$  can differentially weight models  $j$  over the decision space of  $\mathbf{x}$ , incorporating any prior information relevant to model weighting based on past predictive model fit and decision outcomes, and now explicitly allowing

for adjustments based on a currently considered decision  $\mathbf{x}$ . Dependence of  $\pi_j(\mathbf{x})$  on  $\mathbf{x}$  is simply fundamental and critical in our policy setting.

***BPDS calibration functions***

The  $\alpha_j(\mathbf{y}|\mathbf{x})$  are *calibration functions* that define outcome dependence of model weights over the outcome space of  $\mathbf{y}$  for any chosen  $\mathbf{x}$ . This defines the opportunity to increase or decrease the model weights differentially over the outcome  $\mathbf{y}$  space to address model-specific biases and preferences and address questions of model-specific calibration more generally. The BPDS mixture of eqn. (4.1) has the equivalent form

$$f(\mathbf{y}|\mathbf{x}) = \sum_{j=0:J} \tilde{\pi}_j(\mathbf{x}) f_j(\mathbf{y}|\mathbf{x}, \mathcal{M}_j) \quad (4.2)$$

where

$$f_j(\mathbf{y}|\mathbf{x}, \mathcal{M}_j) = \alpha_j(\mathbf{y}|\mathbf{x}) p_j(\mathbf{y}|\mathbf{x}, \mathcal{M}_j) / a_j(\mathbf{x}) \quad \text{and} \quad \tilde{\pi}_j(\mathbf{x}) = k(\mathbf{x}) \pi_j(\mathbf{x}) a_j(\mathbf{x}) \quad (4.3)$$

with normalizing terms  $k(\mathbf{x})$  and  $a_j(\mathbf{x})$  explicitly dependent on  $\mathbf{x}$ . This form shows how the calibration functions  $\alpha_j(\cdot|\cdot)$  modify the initial mixture pdfs  $p_j(\cdot|\cdot) \rightarrow f_j(\cdot|\cdot)$  with corresponding changes of mixture weights  $\pi_j(\mathbf{x}) \rightarrow \tilde{\pi}_j(\mathbf{x})$ .

Note that the choice of relevant calibration functions  $\alpha_j(\mathbf{y}|\mathbf{x})$  will, in any given application, be partly dependent on characteristics of the model pdfs  $p_j(\mathbf{y}|\mathbf{x}, \mathcal{M}_j)$ . In particular, the expectation of each  $\alpha_j(\mathbf{y}|\mathbf{x})$  under  $p_j(\mathbf{y}|\mathbf{x}, \mathcal{M}_j)$  must be finite in order that eqn. (4.1) defines a valid BPDS density  $f(\mathbf{y}|\mathbf{x})$ . Unbounded score functions may sometimes apply, but this point supports the use of bounded scores in general.

***Baseline mixture component***

The model index  $j = 0$  explicitly allows for a *baseline* model component  $\mathcal{M}_0$  in the mixture pdf  $f(\cdot|\cdot)$  that can, among other things, address the ever-present issue of “model set incompleteness” (Tallman and West, 2023, section 2.2.3).  $\mathcal{M}_0$  can be chosen to produce a pdf  $f_0(\cdot|\cdot)$  that is over-dispersed relative to the mixture of the initial  $J$  models, so supporting outcomes  $\mathbf{y}$  that are unusual under the  $J$  models. The baseline

is then a suitable “fall back” model for times when the other models are forecasting poorly.

*Initial mixture*

The special case with each  $\alpha_j(\mathbf{y}|\mathbf{x}) = 1$  defines the *initial mixture* with no BPDS calibration. We use  $p(\mathbf{y}|\mathbf{x})$  in notion, that is,  $p(\mathbf{y}|\mathbf{x}) = \sum_{j=1:J} \pi_j(\mathbf{x})p_j(\mathbf{y}|\mathbf{x}, \mathcal{M}_j)$ .

Special cases fix ideas. First, if  $\pi_j(\mathbf{x}) = \pi_j$  with  $\pi_0 = 0$  are model probabilities based on historical BMA analysis, and with  $\alpha_j(\mathbf{y}|\mathbf{x}) = 1$ , then eqn. (4.1) specializes to BMA. Thus, BMA analyses—with or without this decision dependence in model-specific forecasts—are very special cases of BPDS. Second, again with  $\pi_j(\mathbf{x}) = \pi_j$ ,  $\pi_0 = 0$  and  $\alpha_j(\mathbf{y}|\mathbf{x}) = 1$ , the decision-maker has the freedom to specify the initial mixture probabilities  $\pi_j$  in other ways than with BMA. This includes using historical performance defined by scoring of past forecast outcomes, justifying various approaches to goal-focused model weighting (e.g., Lavine *et al.*, 2021; Loaiza-Maya *et al.*, 2021, and references therein) as special cases of BPDS. Third, mixture BPS (McAlinn and West, 2019; Johnson and West, 2022) is a special case in which models are combined with outcome-dependent weights. In these settings,  $\pi_j(\mathbf{x}) = \pi_j$  depends on past predictive performance,  $p_j(\mathbf{y}|\mathbf{x}) = p_j(\mathbf{y})$  and  $\alpha_j(\mathbf{y}|\mathbf{x}) = \alpha_j(\mathbf{y})$  define outcome-dependent modifications of model probabilities, but there is no decision context so no  $\mathbf{x}$ –dependence. BPDS critically recognizes that the foundational BPS theory allows explicit incorporation of decision goals—admitting the conditioning on  $\mathbf{x}$  throughout all components of eqn. (4.1)—to extend the foregoing analyses.

With predictions of  $(\mathbf{y}|\mathbf{x})$  based on eqn. (4.2), the Bayesian decision-maker acts to identify the optimal decision  $\mathbf{x}$  based on a chosen utility function  $U(\mathbf{y}, \mathbf{x})$ . This involves numerical optimization to maximize the implied expected utility  $\bar{U}(\mathbf{x}) = E_f[U(\mathbf{y}, \mathbf{x})|\mathbf{x}]$  over the decision space of  $\mathbf{x}$ . The notation  $E_f[\cdot|\cdot]$  here explicitly represents expectation with respect to the BPDS distribution, and we use  $E_p[\cdot|\cdot]$  to denote expectation under the initial mixture.

### 4.2.2 Decision-dependent Scores for Calibration Functions

The key step in integrating decision outcomes into relative model weightings is to address the question of how each  $\mathcal{M}_j$  would inform decisions if used alone. Given the predictive pdf  $p_j(\mathbf{y}|\mathbf{x}, \mathcal{M}_j)$  and a chosen, potentially model-specific utility function  $u_j(\mathbf{y}, \mathbf{x})$ , acting based only on  $\mathcal{M}_j$  leads to the optimal decision  $\mathbf{x}_j$  that maximizes  $E_{p_j}[u_j(\mathbf{y}, \mathbf{x})|\mathbf{x}]$  over  $\mathbf{x}$ . The decision-maker has access to this set of model recommendations and is interested in model combination to preferentially weight “good decision models” as well as models that generate good predictions. BPDS formalizes this with specified *score functions*  $\mathbf{s}_j(\mathbf{y}, \mathbf{x}_j)$ , each being a  $k$ -vector of utilities that can be chosen to reflect both predictive and decision goals. The use of multi-dimensional scores addresses multiple goals simultaneously.

The Bayesian decision-theoretic development of Tallman and West (2023) generates the resulting functional forms of the BPDS calibration functions as

$$\alpha_j(\mathbf{y}, \mathbf{x}) = \exp\{\boldsymbol{\tau}(\mathbf{x})' \mathbf{s}_j(\mathbf{y}, \mathbf{x}_j)\}, \quad j = 0:J, \quad (4.4)$$

where  $\boldsymbol{\tau}(\mathbf{x})$  is a  $k$ -vector with elements differentially weighting the multiple utility dimensions of the score vector. The reasoning and theory behind this key result is as follows.

The initial mixture  $p(\mathbf{y}|\mathbf{x}) = \sum_{j=1:J} \pi_j(\mathbf{x}) p_j(\mathbf{y}|\mathbf{x}, \mathcal{M}_j)$  is the  $\mathbf{y}$ -margin of the joint distribution  $p(\mathbf{y}, \mathcal{M}_j|\mathbf{x}) = \pi_j(\mathbf{x}) p_j(\mathbf{y}|\mathbf{x}, \mathcal{M}_j)$ , ( $j = 1:J$ ). Under this initial distribution for any candidate decision  $\mathbf{x}$ , and with score vectors  $\mathbf{s}_j(\mathbf{y}, \mathbf{x}_j)$  defined and evaluated at model-specific optimal decisions  $\mathbf{x}_j$ , the decision-maker has *initial expected score*  $\mathbf{m}_p(\mathbf{x}) = \sum_{j=0:J} \pi_j(\mathbf{x}) \mathbf{m}_{jp}(\mathbf{x})$  where  $\mathbf{m}_{jp}(\mathbf{x}) = \int_{\mathbf{y}} \mathbf{s}_j(\mathbf{y}, \mathbf{x}_j) p_j(\mathbf{y}|\mathbf{x}, \mathcal{M}_j) d\mathbf{y}$ . Treating  $\mathbf{m}_p(\mathbf{x})$  as a benchmark to improve on in expectation, the BPDS theory enquires about distributions  $f(\mathbf{y}, \mathcal{M}_j|\mathbf{x})$  that yield expected scores  $\mathbf{m}_f(\mathbf{x}) \geq \mathbf{m}_p(\mathbf{x}) + \boldsymbol{\epsilon}(\mathbf{x})$  for some non-negative  $k$ -vector (with at least one positive entry)  $\boldsymbol{\epsilon}(\mathbf{x})$ ; this may be chosen to depend on  $\mathbf{x}$ , or may be a specified constant “decision score improvement.” Given  $\mathbf{m}_f(\mathbf{x})$ , the BPDS theory identifies a unique  $f(\cdot, \cdot|\mathbf{x})$  that minimizes the Kullback-Leibler (KL) divergence of  $p(\cdot, \cdot|\mathbf{x})$  from  $f(\cdot, \cdot|\mathbf{x})$  and has an expected score of exactly  $\mathbf{m}_f(\mathbf{x})$ . The

theory is that of *relaxed entropy tilting* (Tallman and West, 2022, 2023; West, 2023) and yields  $f(\mathbf{y}, \mathcal{M}_j | \mathbf{x}) \propto \pi_j(\mathbf{x}) \alpha_j(\mathbf{y}, \mathbf{x}) p_j(\mathbf{y} | \mathbf{x}, \mathcal{M}_j)$  with calibration function precisely as in eqn. (4.4). The *tilting vector*  $\boldsymbol{\tau}(x)$  is implicitly defined by the vector of  $k$  *target score constraints*  $E_f[\mathbf{s}_j(\mathbf{y}, \mathbf{x}_j) | \mathbf{x}] = \mathbf{m}_f(\mathbf{x})$ .

BPDS takes the view that the initial mixture is based on past performance, and additional small changes in models and the way they are weighted based on their expected performance may lead to better future decisions. It asks the question of whether there are perturbations of the mixture based on the initial model probabilities that can lead to improved scores. Consider a stylized example with  $J = 2$  models that in the past have forecast equally well. Traditional model averaging methods focused only on past forecasting experience—and BMA in particular—would confer equal weights in the combination. If, however, the models have different expected scores  $\mathbf{m}_{jp}(\mathbf{x})$ , conferring slightly more weight on the model expected to lead to a higher score makes sense.

The entropic (or exponential) tilting theory is general. It is, of course, possible to tilt the initial joint distribution to most targets, so long as they are technically achievable under the initial distribution. However, an overly ambitious target score will result in a tilted joint distribution that is empirically unreasonable. Hence we emphasize the importance of selecting  $\mathbf{m}_f(\mathbf{x})$  that represents a “small ” improvement over the initial benchmark score  $\mathbf{m}_p(\mathbf{x})$ . This is bolstered by the assumption that the initial model probabilities reflect the empirical plausibility of models, as well as any available information about historical predictive and decision performance. Further, as we exemplify later in the case study, aspects of the computational methodology for model fitting in the sequential time series setting naturally inform on, and allow monitoring of, relevant choices of target expected scores.

### 4.2.3 BPDS Summary

This section has outlined the main ideas underlying BPDS and the key ingredients of the theory and resulting technical machinery. Specifications of score and utility functions, initial model probabilities, and target scores are all required for implementation and are, of course, application specific. The following section develops full details in the context

of the macroeconomic decision-making application. In terms of computation, BPDS requires the use of posterior simulation methods (i.e., draws from conditional predictive densities from each model are required) as well as numerical optimization methods (i.e., to find  $\mathbf{x}_j$  or the overall optimal decision  $\mathbf{x}$  under the final BPDS analysis).

### 4.3 BPDS for Optimal Monetary Policy Decisions

The choice of data and models is inspired by Furlanetto *et al.* (2019). We use quarterly macroeconomic and financial variables from 1973:Q1 to 2022:Q2 from the FRED-QD database maintained by the Federal Reserve Bank of St. Louis. The data set includes GDP (the log of real GDP), prices (the log of the GDP deflator), the interest rate (the shadow rate<sup>1</sup>, which we treat as the policy rate), investment (the ratio of real gross private domestic investment to GDP), stock prices (the log of the S&P500) and the spread (the spread between BAA bonds and the Fed funds rate). Models are run over multiple years. At the end of each quarter, they produce forecasts—full predictive distributions in terms of Monte Carlo samples—of outcomes of interest over the following  $k = 8$  quarters. This is conditional on candidate settings of the decision vector, which is taken as the trajectory of interest (shadow) rates over those quarters. Within-model decision analysis then delivers model-specific optimal decisions about these rates.

#### 4.3.1 Models, Forecasts, and Model-specific Decisions

We consider  $J = 2$  models:  $\mathcal{M}_1$  is a three-variable monetary policy VAR involving GDP, prices, and the interest rate;  $\mathcal{M}_2$  is the model of Furlanetto *et al.* (2019), a VAR with the same variables as  $\mathcal{M}_1$  plus investment, stock prices, and the spread. Following the latter, we include five lags in the VARs. The two structural VARs are identified using the sign restrictions from Table 1 of Furlanetto *et al.* (2019). In  $\mathcal{M}_1$  these restrictions define supply, demand, and monetary policy shocks. In  $\mathcal{M}_2$  investment and financial shocks are additionally identified. We condition on a given value of the policy rate and set the monetary policy to be the driving shock. We do this by imposing restrictions on

---

<sup>1</sup>This is the Federal Funds rate when the latter is positive but can go negative when it is at the zero lower bound, taking into account unconventional monetary policy; see Wu and Xia (2016).

the set of structural shocks underlying the conditional forecasts. Structural shocks other than the monetary policy shock have zero means. We use the asymmetric conjugate prior of Chan (2022), with the advantage that the marginal likelihoods for each can be easily calculated; prior hyperparameter choices are made to maximize the marginal likelihood as in this referenced paper. At each quarter, multi-step ahead predictions are based on simulations using the precision-based sampler of Chan *et al.* (2023). Details on the conditional forecast computations are summarized in Appendix C.2. In short, this generates  $p_j(\mathbf{y}, \mathbf{x})$  with zero-mean constraints on all shocks apart from the monetary policy variable. We do not restrict the variance (i.e., “soft” restrictions) such that we also have uncertainty around the path of  $\mathbf{x}$ . This can be thought of as conditional commitment—we allow the possibility of  $\mathbf{x}$  deviating from the proposed policy path with deviations informed by historical uncertainty around forecasts of the interest rate outcomes.

In this setting, the decision variable is also modelled as an outcome variable. In the current quarter,  $\mathbf{x} = (x_1, \dots, x_k)'$  is the  $k$ -vector of interest rate values over the next  $k$  quarters, and  $\mathbf{y} = (y_1, \dots, y_k)'$  is the corresponding  $k$ -vector of inflation rates. Whatever other variables are in the VAR model  $\mathcal{M}_j$ , our interest focuses on the implied  $p_j(\mathbf{y}|\mathbf{x}, \mathcal{M}_j)$  required for BPDS eqn. (4.1). This conditional predictive is used in decision analysis with the same utility function for each model, namely  $u_j(\mathbf{y}, \mathbf{x}) = U(\mathbf{y}, \mathbf{x})$  given by

$$U(\mathbf{y}, \mathbf{x}) = - \sum_{h=1:k} \{\rho^{k-h}(y_h - y^*)^2 + (x_h - x_{h-1})^2\} \quad (4.5)$$

where  $\rho \in (0, 1)$  is a discount factor. This is a conventional quadratic function that reflects the dual goals of inflation rate targeting and interest rate smoothing over the next  $k = 8$  periods. The  $\mathbf{y}$  terms relate an inflation targeting mandate of  $y^* = 2\%$  over the longer run, while the  $\mathbf{x}$  terms encourage relatively constrained changes in quarter-to-quarter interest rates (the latter being a “don’t rock the boat” consideration, as large swings in interest rates can/will have otherwise unduly effects on the macro-economic system). The terms involving the discount factor  $\rho$  represent the fact that monetary policy works with a lag, so it is desirable to less heavily penalize deviations from the

target at shorter horizons  $h$ . With our two-year horizon ( $k = 8$  quarters), our example analysis below adopts  $\rho = 0.95$ . The model-specific optimal decision vector  $\mathbf{x}_j$  then maximizes  $E_{p_j}[U(\mathbf{y}, \mathbf{x})|\mathbf{x}]$  over  $\mathbf{x}$ . Again, this analysis is repeated each quarter over time, producing rolling updates of the “currently optimal” projections for interest rates over the coming eight quarters.

### 4.3.2 BPDS Model Specification

BPDS requires specification of a relevant class of baseline pdfs  $p_0(\mathbf{y}|\mathbf{x})$ , the model-specific vector score functions  $\mathbf{s}_j(\mathbf{y}, \mathbf{x})$ , the initial BPDS model probabilities  $\pi_j(\mathbf{x})$  as functions of candidate decisions  $\mathbf{x}$ , and the target expected scores  $\mathbf{m}_f(\mathbf{x})$  at any  $\mathbf{x}$ . These are discussed in turn. In addition to customizing BPDS to the specific application, this section highlights a number of methodological developments relevant to other applications, particularly in: (a) the linkages of the  $\pi_j(\mathbf{x})$  to  $\mathbf{x}$  that are relevant more generally when  $\mathbf{x}$  is both an outcome to be forecast as well as a putative decision variable (highlighted in Section 4.3.2 below); and (b) the relevance of dependence structure among the elements of the vector score under the initial distribution  $p(\mathbf{y}, \mathcal{M}_j)$  (highlighted in Section 4.3.2 below).

#### Baseline Distribution

Completing the main BPDS pdf in eqn. (4.1) requires the baseline  $p_0(\mathbf{y}|\mathbf{x})$ . This is taken as a multivariate T distribution with 10 degrees of freedom, using the location from the initial mixture  $p(\mathbf{y}|\mathbf{x})$ , ignoring the baseline (i.e., with  $\pi_0(\mathbf{x}) = 0$ ) and corresponding variance of that mixture inflated by 4. This defines a relevant, tractable  $\mathcal{M}_0$  that can capture outcomes  $\mathbf{y}$  that the two VAR models are not predicting well for any  $\mathbf{x}$  under consideration, and signal that to the decision-maker.

#### BPDS Score Functions

The considerations of inflation targeting and interest rate smoothing reflected in the model-specific decision analysis in Section 4.3.1 are relevant to the choices of BPDS score

functions. Our example takes  $\mathbf{s}_j(\mathbf{y}, \mathbf{x}) = [s_{j1}(y_1, x_1), \dots, s_{jk}(y_k, x_k)]'$  with elements

$$s_{jh}(y_h, x_h) = \exp\{-(y_h - y^*)^2/(2z_y^2)\} + \exp\{-(x_h - x_{h-1})^2/(2z_x^2)\}, \quad h = 1:k, \quad (4.6)$$

where  $y^* = 2\%$  is the inflation target and  $z_y$  and  $z_x$  are *score bandwidth* parameters. This defines a class of bounded score functions, always relevant in decision analysis and here ensuring that the entropically tilted BPDS pdf of eqn. (4.3) is always integrable. The score bandwidths are set so that a certain deviation  $d_y = (y - y^*)^2$  has a score of  $\varepsilon$ ; given a choice of  $\varepsilon$  we set  $z_y = d_y/\sqrt{-2\log(\varepsilon)}$ . Similar considerations apply to choosing  $z_x$ . Our analyses use  $\varepsilon = 0.4$ ,  $d_y = 2$ , and  $d_x = 1$  to ensure the score function is dispersed enough to accommodate modest changes in the Federal Funds rate while being more lenient in deviations from the inflation target. Obvious modifications could incorporate horizon  $h$ -specific inflation targets and differentially weight the two exponential terms, but this form suffices for our main goals in this paper. Note also that, if inflation deviations from target and interest rate changes are “small,” then  $s_{jh}(y_h, x_h)$  is approximately quadratic in  $|y_h - y^*|$  and  $|x_h - x_{h-1}|$  for all  $h$ , perhaps a more familiar utility form.

### Initial Model Probabilities

For clarity in the presentation in this section, we now make explicit the dependency on time, so that the ingredients of the full BPDS predictive pdf in eqn. (4.1)—with the exponential form of the calibration function of eqn. (4.4)—are now indexed by current time  $t$ ; that is,

$$f_t(\mathbf{y}_t|\mathbf{x}_t) \propto \sum_{j=0:J} \pi_{tj}(\mathbf{x}_t) e^{\boldsymbol{\tau}_t(\mathbf{x}_t)' \mathbf{s}_{tj}(\mathbf{y}_t, \mathbf{x}_{tj})} p_{tj}(\mathbf{y}_t|\mathbf{x}_t, \mathcal{M}_j).$$

Bayesian model weighting based on historical predictive performance with respect to defining forecast goals, as developed in (Lavine *et al.*, 2021), provides the starting point for specification of the  $\pi_{tj}(\mathbf{x}_t)$ . The general form adopted is

$$\pi_{tj}(\mathbf{x}_t) \propto \pi_{tj} p_{tj}(\mathbf{x}_t|\mathcal{M}_j), \quad j = 0:J, \quad (4.7)$$

subject to summing to 1 over  $j = 0:J$  and with ingredients as follows.

### *Initial model probabilities*

Traditional Bayesian analysis (e.g., West and Harrison, 1997, chapter 12) defines the starting point. Here the time  $t$  initial model probabilities are based on standard sequential Bayesian updating from those at  $t - 1$ ; that is,  $\pi_{tj} \propto \pi_{t-1,j} p_{tj}(\mathbf{z}_{t-1,j} | \mathcal{M}_j)$  where the “marginal model likelihood” term  $p_{tj}(\mathbf{z}_{t-1,j} | \mathcal{M}_j)$  is the value of the one-step-ahead predictive pdf under  $\mathcal{M}_j$  at the observed values of the last period outcomes  $\mathbf{z}_{t-1,j}$  under that model. In our applied setting, this  $\mathbf{z}_{t-1,j}$  includes time  $t - 1$  outcome values of inflation ( $y$ ), interest rate ( $x$ ), and other economic indicators modelled and forecast in  $\mathcal{M}_j$  in our setting. In general these can differ across models, but in consideration for the initial weights we restrict to variables common across models.

Then, BPDS allows the decision-maker freedom to make alternative choices of the  $\pi_{tj}$ , and the goal and decision focus recommend modification of the standard BMA choice. BMA, after all, only reweights models based on one-step-ahead predictive accuracy. Hence, we adopt two modifications based on recent literature consonant with the goal foci.

First, we use simple power discounting of historically accrued support across models, in which the time  $t - 1$  to time  $t$  evolution is reflected in  $\pi_{tj} \propto \pi_{t-1,j}^\gamma p_{tj}(\mathbf{z}_{t-1,j} | \mathcal{M}_j)$ , where  $\gamma$  is a discount factor in  $(0, 1]$ , closer to 1 for most applications. This acts to discount historically accrued support for model  $j$  at a per-time unit discount rate  $\gamma$  prior to updating by the time  $t - 1$  information. Going back at least to Smith (1979) and then, in a formal dynamic model uncertainty context, West and Harrison (1989, chapter 12, p. 445), power-discounting has been shown to be of value in empirical studies in implicitly allowing for time-variation in the predictive relevance of different models (e.g., Raftery *et al.*, 2010; Koop and Korobilis, 2013; Zhao *et al.*, 2016). Our case study below uses  $\gamma = 0.95$ .

Second, reflecting the foci on specific predictive and decision goals, initial model probabilities should also generally be modified based on the recent relative performance of models with respect to the defined goals. This is the premise underlying the specific

variants of BPS in the setting of adaptive variable selection, or BPS-AVS, in Lavine *et al.* (2021), and related developments in Loaiza-Maya *et al.* (2021), for example. This leads to the immediate BPDS extension of these prior approaches in which the above reasoning is extended to define

$$\pi_{tj} \propto \pi_{t-1,j}^\gamma p_{tj}(\mathbf{z}_{t-1,j} | \mathcal{M}_j) e^{\boldsymbol{\tau}_{t-1}(\mathbf{x}_{t-1})' \mathbf{s}_{t-1,j}(\mathbf{y}_{t-1}, \mathbf{x}_{t-1,j})}.$$

Here the discounted Bayesian model probabilities are further updated with AVS-style weights using the realized BPDS calibration function with relative model scores based on the actual decision outcomes at the last time period. As a result, models are initially and naturally reweighted based on both predictive and decision outcome performance at the last time period.

As a result, in our applied setting, this means that models achieving “good” recent trajectories of interest rate smoothness, as well as relatively accurate forecasting performance of realized inflation outcomes, will be rewarded with higher initial BPDS model probabilities in looking forward to the next time point. And we note that the specification here can cut back to define special cases including BPS-AVS (by setting  $\boldsymbol{\tau}_{t-1}(\mathbf{x}_{t-1}) = \mathbf{0}$ ), and within that to traditional BMA (by setting  $\gamma = 1$ ) for comparisons.

Finally, the inclusion in BPDS of the baseline model and its forecast densities leads to a modification of these initial model probabilities to provide a non-zero value  $\pi_{t0}$  for the baseline. We choose a fixed probability—in our analysis  $\pi_{t0} = 0.1$  at each time point  $t$ —and simply renormalize the  $\pi_{tj}$  above over  $j = 1:J$  accordingly.

### ***Informative conditioning on $\mathbf{x}_t$***

As noted earlier, in our setting the future values of decision variables are also considered outcomes predicted under the models. The models each forecast the future evolution of interest rates as part of the complex, dynamic macroeconomic system, whereas for decisions we must condition on  $\mathbf{x}_t$ . This is reflected in the conditional (on  $\mathbf{x}_t$ ) distributions  $p_{tj}(\mathbf{y}_t | \mathbf{x}_t, \mathcal{M}_j)$  in BPDS where  $\mathbf{x}_t$  is treated as known. The theoretical implication for the BPDS model probabilities is the term  $p_{tj}(\mathbf{x}_t | \mathcal{M}_j)$  in eqn. (4.7)—this is the value

of the current marginal predictive pdf of the vector  $\mathbf{x}_t$  under  $\mathcal{M}_j$ . Assuming the prior (to time  $t$ ) probabilities  $\pi_{tj}$  are specified, this form arises directly via Bayes' theorem. The act of conditioning on  $\mathbf{x}_t$  is informative, and the implied update is, simply by Bayes' theorem, that in eqn. (4.7). Critically, this implies that candidate decision values that are not well-supported under the joint distribution of a model are down-weighted. Conversely, at any candidate decision vector  $\mathbf{x}_t$ , models that are more predictively supportive of the decision  $\mathbf{x}_t$  will be relatively rewarded with higher values of resulting  $\pi_{tj}(\mathbf{x}_t)$ .

In other applications of BPDS, the decision variables may be exogenous, that is, control variables that are to be chosen by the decision-maker but that are not forecast jointly with  $\mathbf{y}_t$  in the set of models. In such cases, it will be common to assume that the external choice of  $\mathbf{x}_t$  is not informative, and then eqn. (4.7) results in decision-independent BPDS probabilities  $\pi_{tj}(\mathbf{x}_t) = \pi_{tj}$  based only on historical data and information.

### BPDS Target Scores

The BPDS target expected score  $\mathbf{m}_f(\mathbf{x}) = E_f[\mathbf{s}(\mathbf{y}, \mathbf{x})]$  represents a desired improvement over the initial expected score  $\mathbf{m}_p(\mathbf{x}) = E_p[\mathbf{s}(\mathbf{y}, \mathbf{x})]$ . In the multi-objective case, the resulting  $\boldsymbol{\tau}(\mathbf{x})$  that defines  $f(\mathbf{y}|\mathbf{x})$  to satisfy this target expectation is sensitive to both the relative scales and dependence of elements of  $\mathbf{s}(\mathbf{y}, \mathbf{x})$  under the initial mixture  $\mathbf{y} \sim p(\mathbf{y}|\mathbf{x})$  at any candidate decision  $\mathbf{x}$ . As functions of  $\mathbf{y}$ , the elements of the random score vector  $\mathbf{s}(\mathbf{y}, \mathbf{x})$  can be strongly correlated, leading to challenges in specifying relevant targets. This can also complicate the calculation of the implied BPDS tilting vector  $\boldsymbol{\tau}(\mathbf{x})$  (i.e., the vector that is needed to satisfy  $\mathbf{m}_f(\mathbf{x}) = E_f[\mathbf{s}(\mathbf{y}, \mathbf{x})]$  under the BPDS density of eqn. (4.1)). We address this by explicitly recognizing score dependencies and defining an approach that explicitly incorporates dependence.

Some theoretical intuition is gained by considered cases of “small perturbations” in which  $\mathbf{m}_f(\mathbf{x}) - \mathbf{m}_p(\mathbf{x})$  has small elements. In this setting, entropic tilting theory in Tallman and West (2022) yields the second-order approximation  $\boldsymbol{\tau}(\mathbf{x}) \approx \mathbf{V}_p(\mathbf{x})^{-1}(\mathbf{m}_f(\mathbf{x}) - \mathbf{m}_p(\mathbf{x}))$  where  $\mathbf{V}_p(\mathbf{x})$  is the variance matrix of  $\mathbf{s}(\mathbf{y}, \mathbf{x})$  under the initial mixture  $p(\mathbf{y}|\mathbf{x})$ . This shows that the implied tilting vector will be very sensitive to the initial score

scales and dependencies as reflected in  $\mathbf{V}_p(\mathbf{x})$ , and suggests a prime focus on a *standardized score* scale; that is, define  $\mathbf{C}_p(\mathbf{x})$  as the scaled eigenvector matrix such that  $\mathbf{V}_p(\mathbf{x}) = \mathbf{C}_p(\mathbf{x})\mathbf{C}_p(\mathbf{x})'$  and set the target score using  $\mathbf{m}_f(\mathbf{x}) = \mathbf{m}_p(\mathbf{x}) + \mathbf{C}_p(\mathbf{x})\boldsymbol{\epsilon}(\mathbf{x})$  for a specified *standardized expected score vector*  $\boldsymbol{\epsilon}(\mathbf{x})$ . The usual convention is taken in which the eigenvector columns of  $\mathbf{C}_p(\mathbf{x})$  are ordered according to decreasingly values of the corresponding eigenvalues, so that the first column is “dominant,” and so forth. This provides insights into how to practically define target scores related to the absolute standardized scale. As examples of the two extremes, taking  $\boldsymbol{\epsilon}(\mathbf{x}) = \epsilon(\mathbf{x})\mathbf{1}$  for some scalar  $\epsilon(\mathbf{x})$  represents targets deviating from the initial expected score in equal amounts of  $\epsilon(\mathbf{x})$  along each of the standardized eigen dimensions. At the other extreme, and most relevant when there are strong score dependencies, taking  $\boldsymbol{\epsilon}(\mathbf{x}) = (\epsilon(\mathbf{x}), 0, \dots, 0)'$  defines the resulting target  $\mathbf{m}_f(\mathbf{x})$  based on the major, dominant eigen dimension alone. The latter is a starting point in general and is taken to define our BPDS case study below. In that setting, we choose  $\epsilon(\mathbf{x})$  such that  $\min\{(\mathbf{m}_f(\mathbf{x})/\mathbf{m}_p(\mathbf{x}))\} = 0.75$  to define the maximum expected improved score in any dimension.<sup>2</sup> It is obviously straightforward to extend this methodology to define target scores impacted by higher eigen dimensions, though that is left for future applications.

### 4.3.3 BPDS Implementation and Optimal Decisions

The final step couples the decision-maker’s utility with the BPDS predictive eqns. (4.1,4.2) to define the optimal decisions from the model synthesis. The decision-maker can adopt any utility function, but an initial neutral analysis will be based on using the same form as usual in the model-specific decisions, the function  $U(\mathbf{y}, \mathbf{x})$  of eqn. (4.5). This is used in the example analysis to follow, with the aim of computing  $\mathbf{x}$  to maximize the implied expected utility function  $\bar{U}(\mathbf{x})$ . In the case study analysis, we compare decisions recommended by BPDS to those from each of the models and to a traditional BMA-based analysis. On the latter, the BMA mixture uses model weights proportional to the marginal likelihoods of the data that are common to all of the models (including in-

---

<sup>2</sup>Additionally, due to the arbitrariness of the signs of eigenvectors, we apply a  $\pm 1$  multiplier to the first column of  $\mathbf{C}_p(\mathbf{x})$  so that the sum of the elements are positive, ensuring the target score improves upon  $\mathbf{m}_p(\mathbf{x})$ .

flation, interest rate, and GDP) under each  $\mathcal{M}_j$ . The BMA mixture naturally involves only the  $p_j(\mathbf{y}|\mathbf{x})$  with no BPS/BPDS outcome-dependence, no notion of a baseline model to address model-set incompleteness, and no regard for the decision-focused use of the models. The foundational BPDS framework theoretically allows for these critical considerations as fundamental to the broader subjective Bayesian decision-analytic and goal-focused approach. Then, technically, BMA arises as a special case of the BPDS analysis as earlier discussed throughout Section 4.3.

The computation of BPDS involves two key components. First, the overall optimization over  $\mathbf{x}$  explores potential BPDS decisions and finds the optimizing vector  $\mathbf{x}$ . This requires an “outer loop” numerical optimization to explore  $\mathbf{x}$  space. In our study, this is performed using a trust region method, namely Powell’s Derivative Free Optimization Solvers (PDFO—Ragoneau and Zhang, 2023). Due to the possibility of multi-modality in  $\bar{U}(\mathbf{x})$ , the optimization is run repeatedly (in parallel) from multiple starting values. In some periods over time, we do find evidence of multi-modality, so repeat starts of the optimization routine are mandated. Second, within each evaluation of a potential BPDS decision, it is necessary to compute the tilting vector  $\boldsymbol{\tau}(\mathbf{x})$  given the constraint  $E_f[\mathbf{s}(\mathbf{y}, \mathbf{x})] = \mathbf{m}_f(\mathbf{x})$  for a target expected score  $\mathbf{m}_f(\mathbf{x})$ . The theoretical basis of this is an implicit equation that is solved via standard, generic numerical optimization methods. Relevant details follow Tallman and West (2023, section 4.4) and are summarized in our Appendix C.1.

The BPDS forecast distributions are evaluated using importance sampling. At any given  $\mathbf{x}$ , the BPDS predictive distribution in eqn. (4.1) is simulated by sampling from the  $p_j(\mathbf{y}|\mathbf{x}, \mathcal{M}_j)$  in proportions defined by the BPDS probabilities  $\pi_j(\mathbf{x})$ . Then, the resulting importance sampling weights are simply proportional to the realized values of  $\alpha_j(\mathbf{y}|\mathbf{x})$ . This provides for efficient computation as well as access to traditional methods and metrics—such as the importance of sampling effective sample sizes (ESS, e.g., Gruber and West, 2016, 2017, in related contexts)—to monitor and evaluate the quality of the resulting Monte Carlo approximations compared to resulting predictive expectations. Note that this evaluation can deliver such metrics to assess “concordance” between the initial densities  $p_j(\mathbf{y}|\mathbf{x}, \mathcal{M}_j)$  and their corresponding BPDS-tilted versions  $f_j(\mathbf{y}|\mathbf{x}, \mathcal{M}_j)$

## Chapter 4. Decision Synthesis in Monetary Policy

in eqn. (4.2), as well as that of the initial mixture  $p(\mathbf{y}|\mathbf{x})$  and the resulting  $f(\mathbf{y}|\mathbf{x})$ . More aggressive BPDS target scores will generally lead to lower concordance, and choices can be partly guided by such empirical evaluations.

## 4.4 Case Study

### 4.4.1 Overview

The analysis and all empirical results proceed sequentially on an expanding window of data beginning in 1992Q2. Our summaries begin with a comparison of the decisions recommended by BPDS to those suggested by BMA. This is followed by a discussion of the individual models and how they are combined by BPDS and BMA. Additional discussion highlights some operational BPDS details to provide further insights into the resulting decision outcomes.

### 4.4.2 Optimal Decisions

Figure 4.1 shows the actual policy rate each quarter along with the 1–8 quarters-ahead policy recommendations that would have been made by BPDS and BMA. In using the shadow rate, the zero lower bound is not in effect and negative values for the policy rate are possible. Recommendations for negative values for the policy rate are not to be taken literally as advising cuts to a negative Federal Funds rate, but rather as a suggestion to undertake other forms of monetary easing that would be expected to proxy such cuts.

Since 2014, the optimal policy paths recommended by the two approaches are generally similar, though there are notable differences prior to that time. Some specific periods of interest are now highlighted.

#### *2014 to the present*

During this period, BMA and BPDS provide similar recommendations that are often quite different from the actual policy rate. For almost all of these times, the policy recommendations are to cut interest rates, whereas (apart from 2019–2021) the actual policy rate increased. Some differences do arise between BMA and BPDS. For example, during the post-COVID inflation period, BPDS recommends a higher rate path. BPDS is closer to the decision actually made by the Federal Reserve, although according to BPDS interest rates should decrease throughout 2023 and 2024.

*The financial crisis and subsequent recession*

It is during this period that the differences between BPDS and BMA are most acute. The actual policy rate fell slowly during this period. BPDS recommends rate cuts as well, initially at a more rapid rate than what actually occurred, but as of 2010, its recommendations are similar to the ones the policymakers actually made. In contrast, BMA recommends huge cuts to the policy rate right at the start of the financial crisis, but subsequently consistently argues for rate increases.

*The first years of the 21st century*

From 2003 through to the beginning of the financial crisis, the actual policy rate was gradually increasing. In this period, BMA consistently recommends rapid rate increases. In contrast, BPDS recommendations are generally similar to what actually transpired, apart from at the beginning of this period where the advice is to raise the policy rate more slowly than what actually occurred.

*The 1990s*

During this period, the pattern is more mixed. Optimal policy recommendations generated under each of BMA and BPDS often differ from actual decisions, with no consistent pattern; at times the recommended rates are higher than the actual policy rate, and at other times they are lower.

A general pattern, one that occurs throughout the sample period, is that BMA and BPDS typically recommend larger changes in policy rates than were actually implemented by policymakers. Part of this is presumably due to differences between the policymakers' utility function and those used in our analyses. Other possible explanations are that policymakers can affect expectations through their communications, which is a channel not captured in the model, or their models use a much steeper Phillips Curve. Also, we focus only on inflation up to two years ahead, without considering the possibility of an over- or under-shoot of inflation after eight quarters. In contrast, policymakers would generally aim for inflation to be sustainably at target over the longer-term.

### 4.4.3 Trajectories of BPDS and BMA Predictive Densities

Figures 4.2–4.4 shed light on these patterns. These images represent the time trajectories of predictive densities of inflation at each relevant horizon, using BPDS (Fig. 4.2) and BMA (Fig. 4.3), as well as their differences (Fig. 4.4). These indicate that the BPDS mixtures are less dispersed than the BMA mixture for much of the sample period; that is, BMA predictive distributions are relatively more heavy-tailed, especially at longer horizons. This is partly due to the BPDS score function emphasizing that the policy-maker wants to avoid extreme inflation outcomes, and also accounts for why BMA often tells policymakers to make larger changes to the policy rate than BPDS, as discussed in Section 4.4.2. The differences between BPDS and BMA become larger at longer forecast horizons. Medium- and long-term macroeconomic forecasting is difficult, which leads to standard methods such as BMA producing fairly dispersed predictive densities at longer horizons. BPDS, on the other hand, is reducing this effect, which dampens the BPDS optimal decisions and reduces predictive uncertainty relative to BMA. Then, differences between BPDS and BMA forecast densities are reduced after the financial

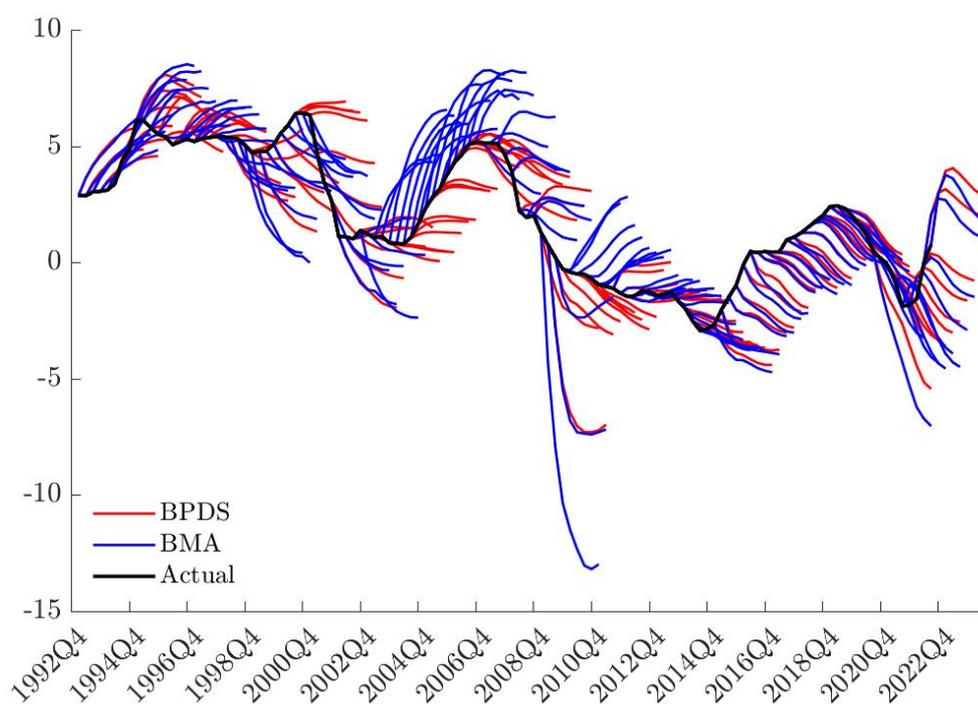


Figure 4.1: Recursively calculated policy decisions

crisis, which helps account for why their policy recommendations are similar in the last decade of the sample.

#### 4.4.4 Model Probabilities

Figure 4.5 shows the trajectories of model probabilities under BPDS and BMA. These are the discounted AVS prior model probabilities  $\pi_{tj}$  over time, the implied initial decision-dependent probabilities  $\pi_{tj}(\mathbf{x}_t)$  evaluated at the BPDS-optimal decision  $\mathbf{x}_t$  at each time, the resulting BPDS probabilities  $\tilde{\pi}_{tj}(\mathbf{x}_t)$  of eqn. (4.3) at each time, and the standard BMA probabilities over time.

Under traditional BMA, the two model probabilities are appreciable until the financial crisis. After the start of the crisis, the less parsimonious  $\mathcal{M}_2$ , which includes additional financial variables, receives virtually all the weight. In contrast, BPDS weights vary more over time, allocating most of the weight to the parsimonious  $\mathcal{M}_1$  for much of the period (i.e., 1997 through 2017), though  $\mathcal{M}_2$  plays more of a role at both the beginning and end of the sample period. That BPDS generally favours the more parsimonious  $\mathcal{M}_1$ , with less dispersed forecast distributions, partially accounts for why BPDS often dampens extreme recommendations made when using BMA.

BPDS probabilities on the over-dispersed  $\mathcal{M}_0$  are generally small, though with notable increases at two critical periods: the start of the financial crisis and the start of the COVID-19 pandemic. In such extreme times, when neither  $\mathcal{M}_1$  nor  $\mathcal{M}_2$  forecasts well, the increased probability on the fall-back  $\mathcal{M}_0$ —though small—provides an indicator of this.

The BPDS prior model probabilities  $\pi_{tj}$  based on discounted AVS differ noticeably from BMA probabilities (except at the very start of the time period). A big impact then arises from the conditioning on information generated in the decision space to map these  $\pi_{tj}$  to the decision-dependent weights  $\pi_{tj}(\mathbf{x}_t)$  at the BPDS optimal decisions  $\mathbf{x}_t$  at each time. Recall that this mapping theoretically properly takes into account the likelihood of future, as yet unobserved, interest rate outcomes; this relevant information is not accounted for in the prior weights  $\pi_{tj}$  and is, of course, absent under BMA. The subsequent map from initial probabilities  $\pi_{tj}(\mathbf{x}_t)$  to the BPDS weights  $\tilde{\pi}_{tj}(\mathbf{x}_t)$  is

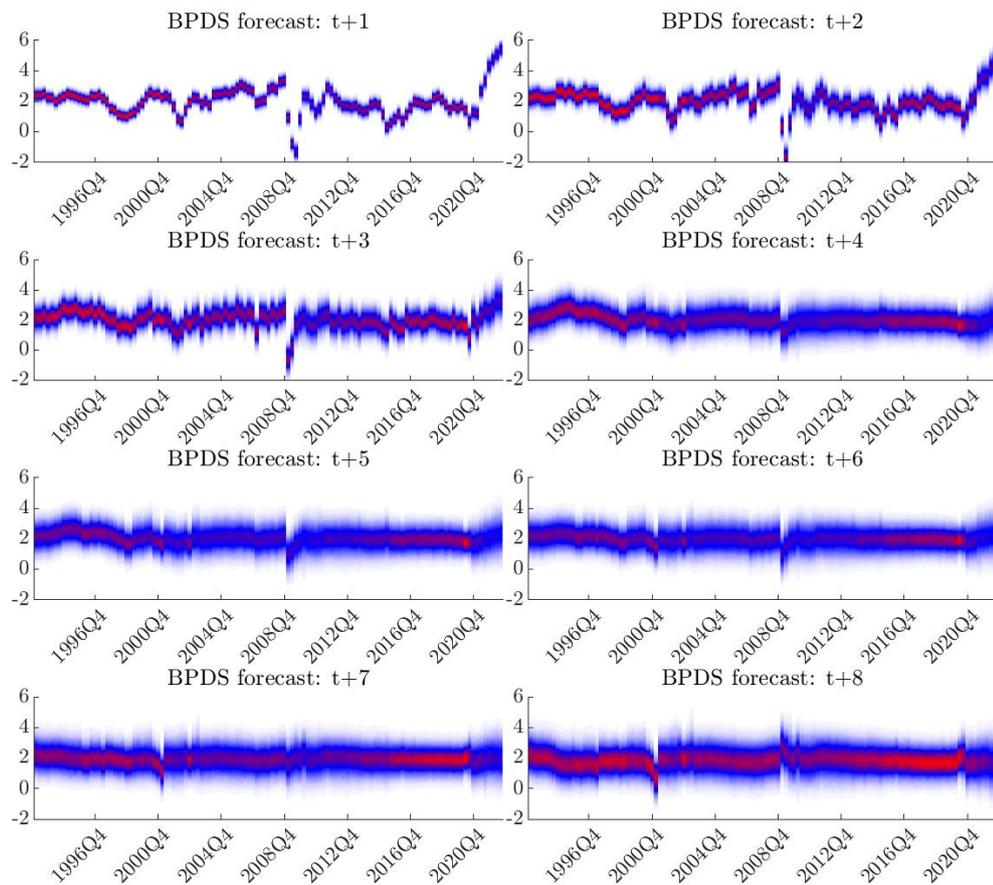


Figure 4.2: BPDS forecast densities of inflation. The frames represent 1–8 quarter ahead forecasts, reading along the rows from top-left to bottom-right. The colours represent probabilities, with the blue shading showing lower probability, and red showing higher probability.

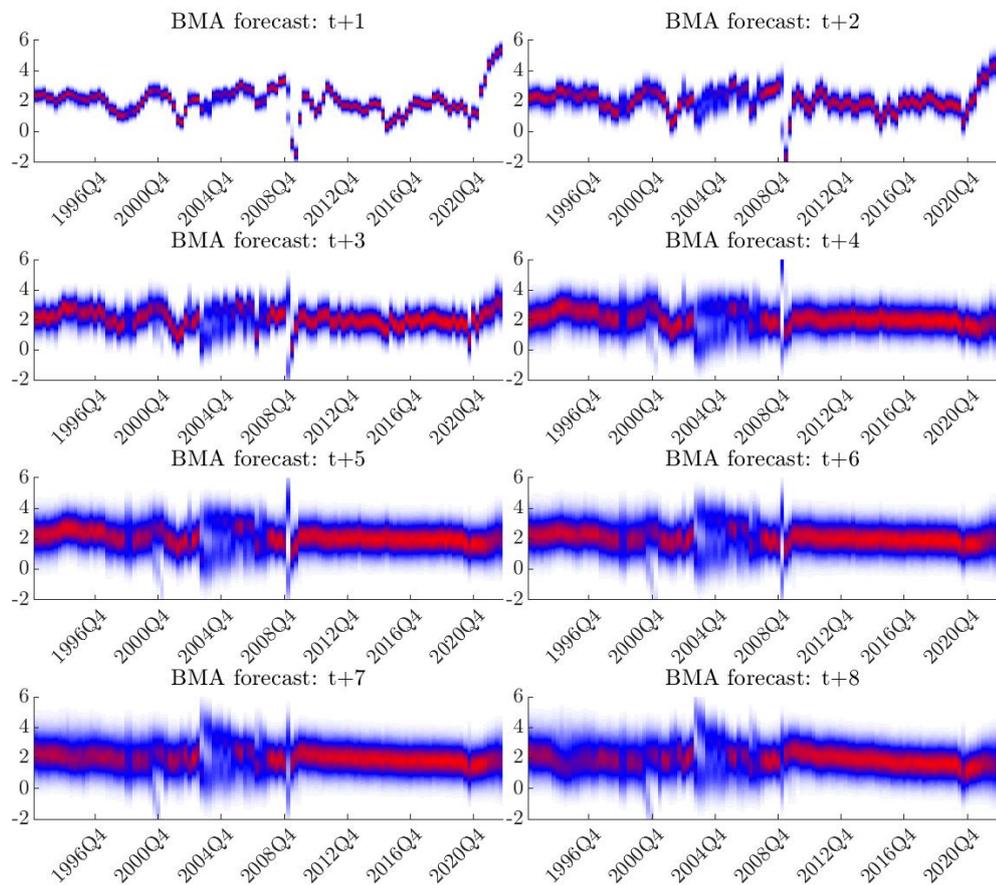


Figure 4.3: BMA forecast densities of inflation. The frames represent 1–8 quarter ahead forecasts, reading along the rows from top-left to bottom-right. The colours represent probabilities, with the blue shading showing lower probability, and red showing higher probability.

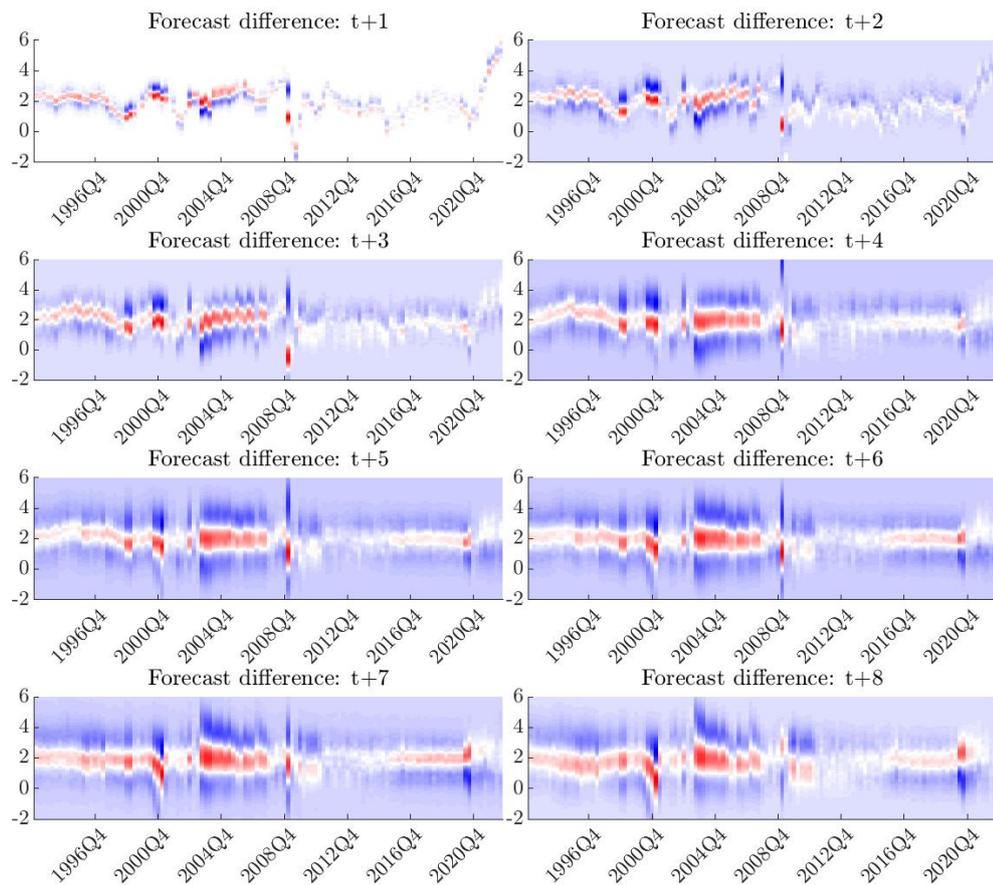


Figure 4.4: Difference between BPDS and BMA forecast densities of inflation. Red-shaded regions have higher probability under BPDS than under BMA, with blue shading indicating the reverse, and white shading showing equal probability.

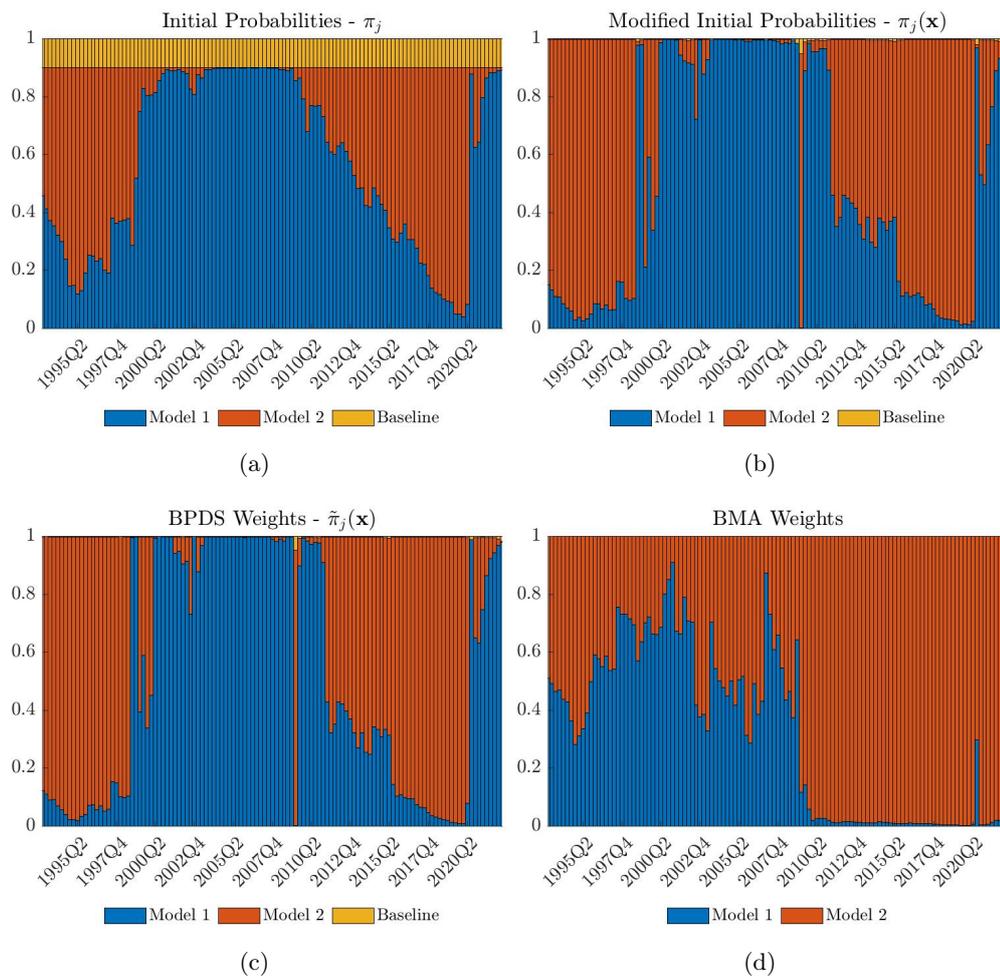


Figure 4.5: Time trajectories of model probabilities. (a) Prior BPDS probabilities  $\pi_{tj}$  based on discounted AVS with a fixed baseline  $\pi_{t0} = 0.1$ ; (b) BPDS decision-dependent initial probabilities  $\pi_{tj}(\mathbf{x}_t)$ ; (c) Implied BPDS weights  $\tilde{\pi}_{tj}(\mathbf{x}_t)$ ; (d) BMA probabilities.

wholly based on the impact of the entropic tilting towards “more favourable” decisions, in expectation. We see that the impact is rather small over time, and this is to be expected: the BPDS analysis uses “small” perturbations of the initial mixture based on target expected scores that are only modest increases over those under the initial mixture. We expect to see slight tilting towards models that are expected to do well, but not large changes relative to the initial probabilities.

### 4.4.5 Additional Insights from BPDS Results

Time trajectories of the evaluated tilting vectors  $\tau_t(\mathbf{x}_t)$ , evaluated at the optimal decisions  $\mathbf{x}_t$ , are shown in Figure 4.6. The values generally tend to increase with horizon  $h$ , thus attaching more weight to longer forecasting horizons. This is partly to be expected due to the higher uncertainties at longer forecast horizons.

Figure 4.7 plots trajectories of several effective sample size (ESS) measures arising from the importance sampling to simulate BPDS predictive distributions, as discussed in

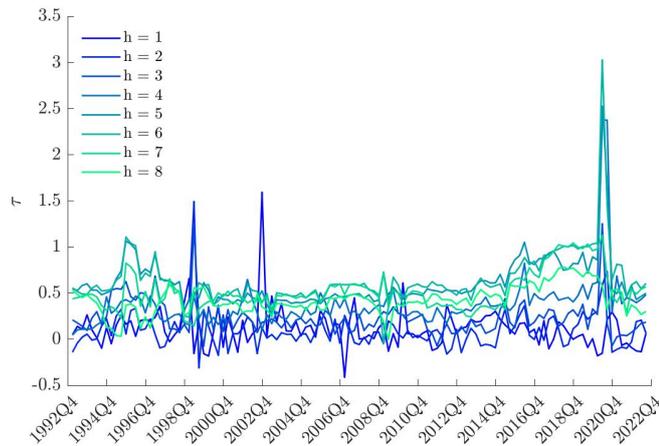


Figure 4.6: Trajectories of the eight elements of the evaluated BPDS tilting vector  $\tau_t(\mathbf{x}_t)$  at the optimized  $\mathbf{x}_t$  at each quarter.

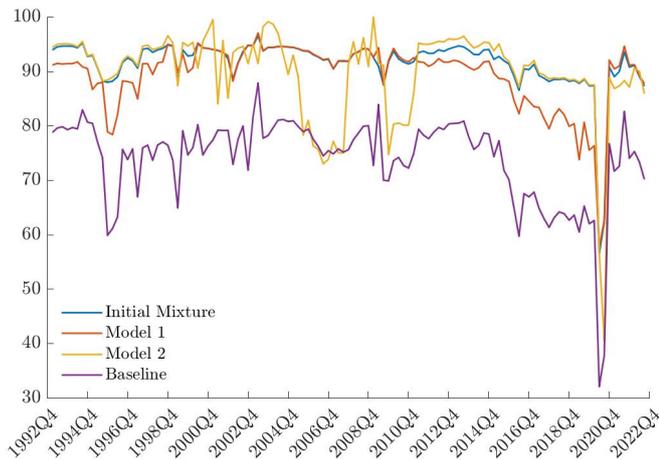


Figure 4.7: Trajectories of the effective sample size (ESS) metrics for individual models and for the BPDS initial mixture.

Section 4.3.3. This provides a read-out of the extent of tilting the initial mixture  $p(\mathbf{y}|\mathbf{x})$  to the BPDS mixture  $f(\mathbf{y}|\mathbf{x})$ , as well as that for tilting each of the individual model pdfs from  $p_j(\mathbf{y}|\mathbf{x}, \mathcal{M}_j)$  to  $f_j(\mathbf{y}|\mathbf{x}, \mathcal{M}_j)$  (again, time-indexed and updated throughout the time series). Until the COVID recession, the ESS of the initial mixture is stable between 90–95%, suggesting only a small amount of tilting, as desired. The COVID recession is a period of rapid change, as expected, as we see large changes in the initial weights and larger values of  $\tau$  required to achieve the desired target. The low value of ESS indicates that at that point, the target expected scores are unrealistic given the then-current state of the economy. However, the resulting decisions during this time period appear to be rather sensible. This means we do not need to be too concerned about the low ESS, which can, in any case, be redressed by simply increasing the overall Monte Carlo sample size accordingly. The ESS values of individual models are generally lower than that of the overall mixture and somewhat more volatile. One nice point is that, even when one of the models seems to suffer a low ESS, the BPDS mixture ESS is generally maintained at higher values. This indicates that BPDS is able to strike a balance in weighting expected versus historical performance of models on both predictive and decision outcomes.

Finally, Figure 4.8 compares the realized trajectories of expected utilities under BPDS and BMA. Each uses the same utility function to define the final optimal policy path decision, so these are directly comparable, and the comparison is relevant in terms of the setting of forward, sequential decisions where a change to much lower values at any time point should signal concern to the decision-maker. BPDS is designed to target an expected utility higher than that of the initial mixture, but whether it achieves a higher expected utility than BMA—which has different initial probabilities and lacks outcome-dependent weighting—is a question for empirical study. In this example, as illustrated in the figure, BPDS utility does exceed that of BMA in virtually every period. After the financial crisis, the two are similar, consistent with the earlier finding that they typically produced similar decisions during this time period. However, before the financial crisis, there were several periods during which the BPDS expected utilities were substantially higher than those of BMA. These correspond to times where we see more

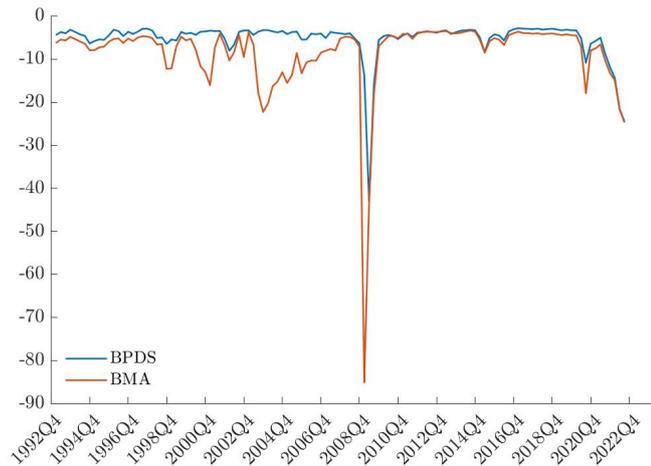


Figure 4.8: Trajectories of expected utilities comparing BPDS with BMA.

differences between optimal policy path recommendations. Within these time there are some periods of greater concordance between BPDS and actual policy decisions, as well as more constrained (i.e., less extreme) recommended decisions under BPDS relative to BMA.

## 4.5 Summary Comments

BPDS is the formal, foundational Bayesian framework that extends traditional Bayesian model uncertainty analysis to address explicit use of model-specific decision outcomes as well as purely predictive performance in model comparison and combination. This paper has adapted the BPDS foundations to define implied methodology in formulating macroeconomic decision-making when faced with multiple objectives and multiple outcomes of interest in the monetary policy setting.

Earlier applications of BPDS have focused mainly on financial portfolio forecasting and decisions (Tallman and West, 2023, section 6; Tallman and West, 2024). In this setting, forecasting models do not (generally) depend on the decisions of interest, while utility functions may and often do depend on the models and their predictions. In contrast, the setting of monetary policy analysis is one in which the dependence of models and their forecasts on the decision variables (policy instruments) is simply fundamental. It is also a setting in which the decision variables are treated simultaneously as

outcomes. The future paths of central bank interest rates, for example, are modelled as time series outcomes along with other economic and financial indicators in VAR models. This then leads to conditioning on decision variables to define predictions of other indicators, with consequent implications for relative model weights in the model uncertainty setting. This latter point is critical as it then leads to relatively up- or down-weighting a model based on how well-supported a particular candidate decision is under its predictions; to our knowledge, this is the first time this central question has been formally, statistically addressed. These central features of predictive decision-making in monetary policy contexts are addressed with extensions and customization of the existing theory of BPDS.

The BPDS perspective—of integrating historical and expected decision outcomes with focused aspects of statistical predictive performance into relative model weightings—is new to the policy arena. We argue for this perspective since policymakers are primarily interested in using sets of models for the eventual policy decisions. Pure forecasting exercises—and evaluation and combinations of models for prediction *per se*—are, of course, of parallel interest and importance. We emphasize that BPDS also involves addressing predictive performance on specific, defined outcomes of interest. But most importantly, by putting the spotlight on decision-making, we gain additional insights into policy-making that are not possible in exercises that focus solely on predictive performance.

In a recursive, real-time decision-making exercise, we find substantial differences at various periods of time between the policy recommendations of BPDS and the traditional Bayesian model averaging approach, though good concordance at other times. When recommended policy decisions differ between the approaches, in most cases the BPDS policy paths are more intuitively sensible and less extreme than under BMA, and more consistent with the actual decisions made by the policymakers at the time. The case study presented investigates and interprets aspects of BPDS in terms of differential model weights based on historical information alone, and then updated based on identified optimal decisions, with consequent insights into how the differences relative to standard BMA arise and are exploited. This case study is a first step towards broader

## Chapter 4. Decision Synthesis in Monetary Policy

development and evaluation of BPDS in a setting with larger numbers of econometric models. The parallel next steps will naturally include BPDS for scenario forecasting. These analyses, in collaboration with policymakers, explore and aim to understand the sensitivity of model-based recommendations relative to chosen potential economic scenarios.

# Chapter 5

## Conclusion

### 5.1 Summary Comments

Policy makers often use multiple models to hedge against the weaknesses of individual macro-economic models. In this context it is important to understand how best to utilize multiple models both in terms of using them for decision-making and forecasting. This thesis explores several issues related to the use of multiple models in macro-economic policy making.

In 2, I investigate different approaches for combining large numbers of density predictions. This is an important issue since many practical applications can involve large numbers of forecasts, such as nowcasting systems or combining survey forecasts. I use two common approaches in economics to deal with large datasets: global-local shrinkage priors and factor modeling. In particular, I use the newly developed triple gamma prior, and the priors it encompasses, along with a novel factor modeling approach to density combinations.

I test the approaches in two very different applications: a model-based nowcasting exercise on Canadian real GDP, and forecasting Euro Area real GDP growth using distributions from the Survey of Professional Forecasters. These two applications cover two regions, have different forecast horizons, include model-based and survey-based predictions, and the evaluation sample covers the Great Financial Crisis, Euro Area Crisis, and COVID-19 pandemic, allowing for a comprehensive assessment of the various

synthesis functions. First, I find that constant parameter specifications tend to perform better than their time-varying counterparts. This shows that in applications with little structural change, relatively short samples, and a large cross-section of models, a more parsimonious model is preferable. This is an important finding as recently developed combination schemes tend to utilize time-varying parameter specifications. Second, and more importantly, I find that shrinkage approaches generally outperform factor-model-based combinations.

It is interesting to note that the two synthesis functions imply very different weighting structures. The sparse weighting scheme of shrinkage priors implies that decision-makers should give considerable weight to a smaller set of experts. This, however, carries the risk of “putting all your eggs in one basket,” which at times could adversely affect the performance of the sparse combinations. In contrast, the factor-model-based combination implies a dense weighting scheme, which produces a “consensus” forecast. Overall, my results suggest that focusing on a parsimonious combination that considers a smaller set of accurate experts is preferable to following the herd.

Chapter 3, extends the previous work by examining nonparametric density forecast combination using regression trees: BPS-RT. While a handful of papers use nonparametric techniques to combine densities, ours is the first to use regression trees. We model the combination weights using the regression trees which keeps the model linear in parameters such that interpretation is easier. Regression trees use covariates, or weight modifiers, to drive changes in parameters which allows us to explain the combination weights. Which is in contrast to conventional BPS applications where model parameters follow a random walk. Taken together, our approach is flexible but retains interpretability through linearity and the use of weight modifiers.

We test the performance of BPS-RT in two different applications – combining model-based US inflation density forecasts and subjective histogram-based forecasts of euro area GDP growth. We find that, across both applications, BPS-RT forecasts well in terms of both relative and absolute accuracy. Interestingly, and in contrast to standard BART applications, we find that using a parsimonious single-tree specification outperforms models with more trees. Which is thematically similar from the findings in

## Chapter 5. Conclusion

Chapter 3 which show simpler models perform better. Inspecting the best-performing specification, we observe that this superior performance is due to less disperse forecast densities and BPS-RT's ability to better accommodate the shocks associated with the global financial crisis (in the GDP application) and COVID-19 (in the inflation application). Our proposed measure of model set incompleteness suggests that BPS-RT is able to capture much of the post-COVID rise in inflation. Triggered by a rise in the relative importance of the time trend in determining tree splits, itself highlighting the unusual nature of this inflationary period, BPS-RT also shifts its combination weights toward component models with SV. This contrasts with the prior period of lower inflation, when the business cycle indicators were found to be the more important weight modifiers.

Chapter 4 shows how to use multiple models to make decisions by using BPDS for monetary policy decision-making. We build on earlier applications of BPDS that have focused mainly on financial portfolio forecasting and decisions (Tallman and West, 2023, section 6; Tallman and West, 2024). In portfolio settings, the forecast typically do not depend on the decisions of interest – individual agents can rarely effect the stock market value. In contrast, monetary policy decisions are made with the intent of affecting inflation. This creates a challenging environment since we weight models by outcomes which are then a function of the decision. Therefore the weights are a function of both the outcome and decision. There are other aspects of monetary policy decision-making contexts that add to the challenge. For example, monetary policy is made for outcomes over multiple horizons, there can be multiple objectives, and the decision variables are also endogenous to the models. This latter point is critical as it then leads to relatively up- or down-weighting a model based on how well-supported a particular candidate decision is under its predictions; to our knowledge, this is the first time this central question has been formally, statistically addressed. Despite the challenges we develop a BPDS approach that takes into account these aspects of the monetary policy decision-making context. Specifically, we weight forecasts at each forecast horizon, weight models by expected and realized decision criteria, their empirical fit, and by the plausibility of a proposed policy path.

This BPDS perspective—of integrating historical and expected decision outcomes with focused aspects of statistical predictive performance into relative model weightings—is new to monetary policy decision making. And we argue that this approach is preferable to classical approaches based on forecasting accuracy. This is because policymakers are primarily interested in using sets of models for the eventual policy decisions. This is not to diminish the importance of pure forecasting exercises—and evaluation and combinations of models for prediction *per se*. We would be remiss not to emphasize that BPDS also involves addressing predictive performance on specific, defined outcomes of interest. But by focusing on decision-making, we gain additional insights into policy-making that are not possible in exercises that focus solely on predictive performance.

In a recursive, pseudo real-time decision-making exercise, we find substantial differences at various periods of time between the policy recommendations of BPDS and the traditional Bayesian model averaging approach, though high concordance at other times. When recommended policy decisions differ between the approaches, in most cases the BPDS policy paths are more sensible and less extreme than under BMA, and more consistent with the actual decisions made by the policymakers at the time. The case study presented investigates and interprets aspects of BPDS in terms of differential model weights based on historical information alone, and then updated based on identified optimal decisions, with consequent insights into how the differences relative to standard BMA arise and are exploited.

Overall this thesis explored issues related to using multiple models for forecasting and decision-making. I showed how to combine forecasts when faced with large numbers of models finding that sparse combinations techniques perform better. This implies it is better to choose a handful of models than average of a large number of models. Second, I show how tree based synthesis functions could be useful for forecast combination. Namely, that they offer some advantages in being able to explain the combination weights. Finally, I show how to use multiple model in monetary policy decision-making employing BPDS in a novel and challenging application.

## 5.2 Further Research

Future lines of research could involve further development of synthesis functions for forecast combination. Otherwise using tree-based synthesis functions in other forecasting applications and contexts could be fruitful. Specifically, we could investigate the usefulness of different sets of weight modifiers and the implications for weight structure. For instance, this could draw on the ability of BPS-RT, via its choice of weight modifiers, to capture general patterns of cross-sectional dependence between competing agents' probabilistic forecasts. Additional structure could be given to the clustering by, for example, letting the combination weight on a given individual agent's density forecast depend not only on characteristics of her own forecast (such as its mean or variance) but on characteristics of the other agents' forecasts. There could also be further work on the broader development and evaluation of BPDS in monetary policy makings. These could be settings with more diverse set of econometric models or different monetary policy regimes or decision-maker utility functions. For example, incorporating preferences for financial stability or to 'lean against the wind' when there are high debt levels. Further developments could include BPDS for scenario forecasting. These analyses, in collaboration with policymakers, aim to understand the sensitivity of model-based recommendations relative to chosen potential economic scenarios.

# Appendix A

## Chapter 2 Appendix

### A.1 Tables

Table A.1: Nowcasting Application: Overview of Forecasting Performance

	Global-Local Shrinkage Priors										Factor Model Combinations									
	Time-Varying					Constant					Time-Varying					Constant				
	DLM	Lasso	DG	TG	HS	Lasso	DG	TG	HS	1 Factor	2 Factor	3 Factor	4 Factor	5 Factor	1 Factor	2 Factor	3 Factor	4 Factor	5 Factor	
24 weeks until NA	2.77	2.73	2.67	2.62	2.63	3.30	2.68	2.67	2.69	2.58	2.57	2.58	2.58	2.57	2.87	2.77	2.76	2.76	2.76	
22 weeks until NA	2.74	2.67	2.60	2.59	2.58	3.22	2.64	2.77	2.67	2.57	2.58	2.57	2.58	2.57	2.84	2.77	2.75	2.74	2.74	
20 weeks until NA	2.20	2.14	2.02	2.06	2.02	2.53	1.92	1.91	1.92	2.58	2.55	2.55	2.55	2.54	2.91	2.02	2.03	2.03	2.03	
18 weeks until NA	2.28	2.08	1.94	1.86	1.99	2.43	1.97	1.87	1.91	2.53	2.51	2.49	2.50	2.49	2.66	1.93	1.91	1.90	1.91	
16 weeks until NA	1.72	1.83	1.71	1.55	1.60	2.12	1.64	1.53	1.54	2.19	2.23	2.24	2.21	2.20	1.93	1.86	1.83	1.82	1.80	
14 weeks until NA	1.59	1.74	1.66	1.53	1.53	2.01	1.57	1.61	1.57	2.01	2.11	2.11	2.09	2.05	1.83	1.76	1.73	1.72	1.70	
12 weeks until NA	1.31	1.46	1.29	1.14	1.14	1.69	1.21	1.14	1.15	1.34	1.36	1.36	1.36	1.35	1.27	1.26	1.24	1.24	1.23	
10 weeks until NA	1.30	1.47	1.29	1.09	1.12	1.67	1.20	1.11	1.20	1.33	1.35	1.34	1.34	1.33	1.26	1.24	1.22	1.21	1.21	
8 weeks until NA	0.89	0.96	0.92	0.83	0.82	1.19	0.82	0.79	0.81	1.30	1.28	1.26	1.25	1.24	1.35	1.06	1.07	1.06	1.05	
6 weeks until NA	0.85	0.94	0.88	0.78	0.80	1.16	0.80	0.77	0.79	1.29	1.24	1.23	1.22	1.21	1.33	1.02	1.02	1.01	0.99	
4 weeks until NA	0.58	0.68	0.62	0.57	0.56	0.80	0.54	0.54	0.54	0.58	0.58	0.58	0.58	0.58	0.56	0.53	0.55	0.55	0.55	
2 weeks until NA	0.58	0.68	0.61	0.56	0.56	0.80	0.54	0.54	0.54	0.58	0.58	0.57	0.58	0.58	0.55	0.53	0.56	0.56	0.56	

Notes: The rows show prediction horizons in weeks until the release of the National Accounts (NA). Periods 24 and 22 weeks until the National Accounts are the forecast periods, while 20 to 10 weeks is the nowcast period, and 8 until 2 weeks is the backcast period. The columns correspond to the Dynamic Linear Model benchmark (DLM), constant and time-varying specification of the Lasso, double gamma prior (DG), triple gamma prior (TG), Horseshoe prior (HS) and factor model synthesis functions with 1 to 5 factors.

Table A.2: SPF Application: Overview of Forecasting Performance

	Global-Local Shrinkage Priors										Factor Model Combinations									
	Time-Varying					Constant					Time-Varying					Constant				
	DLM	Lasso	DG	TG	HS	Lasso	DG	TG	HS	1 Factor	2 Factor	3 Factor	4 Factor	5 Factor	1 Factor	2 Factor	3 Factor	4 Factor	5 Factor	
Wide dataset	1.33	1.73	1.47	1.56	1.50	1.34	1.32	1.37	1.36	1.67	1.65	1.64	1.64	1.62	1.58	1.57	1.54	1.54	1.51	
Tall dataset	1.32	1.47	1.51	1.54	1.45	1.24	1.23	1.22	1.22	1.63	1.61	1.62	1.60	1.61	1.19	1.17	1.21	1.23	1.18	

Table A.3: The rows show results for wide and tall datasets. The columns correspond to the Dynamic Linear Model benchmark (DLM), constant and time-varying specification of the Lasso, double gamma prior (DG), triple gamma prior (TG), Horseshoe prior (HS), and factor model synthesis functions with 1 to 5 factors.

## A.2 Technical Appendix

### A.2.1 MCMC Algorithm

This section describes the Markov Chain Monte Carlo (MCMC) algorithm used to estimate the forecast combinations. It largely follows McAlinn and West (2019) for the BPS steps, Cadonna *et al.* (2020) for the global-local shrinkage priors combinations, and Lopes and West (2004) for the factor model combinations. The MCMC follows a two-component block Gibbs sampler: one component samples the synthesis function parameters, and the second samples from the expert forecast distributions or the agent states. As such, I discuss the estimation of each synthesis function separately, followed by details on sampling the agent states.

### A.2.2 Global-local Shrinkage Combinations

This section describes the estimation of the global-local shrinkage synthesis functions. Knaus *et al.* (2021) provide an R package and the vignette is an excellent overview of the estimation and priors of these models. More details are available in Cadonna *et al.* (2020) and Bitto and Frühwirth-Schnatter (2019). I first describe the model, followed by the priors, and then describe the MCMC algorithm.

Starting with the centered parameterization of the synthesis function, for  $t = 1, \dots, T$ , we have that

$$y_t = x_t \beta_t + \epsilon_t \quad \beta_t = \beta_{t-1} + u_t \epsilon_t \sim \mathcal{N}(0, \sigma_t^2) \quad u_t \sim \mathcal{N}(0, Q) \quad (\text{A.1})$$

where  $y_t$  is a univariate response variable and  $x_t = (x_{t0}, x_{t1}, \dots, x_{td})$  is a  $d$ -dimensional row vector containing the regressors at time  $t$ , with  $x_{t1}$  corresponding to the intercept.

For simplicity, I assume here that  $Q = \text{Diag}(\theta_1, \dots, \theta_d)$  is a diagonal matrix, implying that the state innovations are conditionally independent. Moreover, I assume the initial value follows a normal distribution (i.e.,  $\beta_0 \sim \mathcal{N}_d(\beta, Q)$ ), with initial mean  $\beta = (\beta_1, \dots, \beta_d)$ . Model (A.1) can be rewritten equivalently in the non-centered parametrization as

$$\begin{aligned} y_t &= x_t \beta + x_t \text{Diag}(\sqrt{\theta_1}, \dots, \sqrt{\theta_d}) \tilde{\beta}_t + \epsilon_t, & \epsilon_t &\sim \mathcal{N}(0, \sigma_t^2) \\ \tilde{\beta}_t &= \tilde{\beta}_{t-1} + \tilde{u}_t, & \tilde{u}_t &\sim \mathcal{N}_d(0, I_d) \end{aligned} \quad (\text{A.2})$$

with  $\tilde{\beta}_0 \sim \mathcal{N}_d(0, I_d)$ , where  $I_d$  is the  $d$ -dimensional identity matrix. Furthermore, the model can accommodate stochastic volatility or constant volatility. In the former case, the log-volatility  $h_t = \log \sigma_t^2$  follows a random-walk. More specifically,

$$h_t | h_{t-1}, \sigma_\eta^2 \sim \mathcal{N}(h_{t-1}, \sigma_\eta^2), \quad (\text{A.3})$$

with initial state  $h_0 \sim \mathcal{N}(a_0, b_0)$ .

### Shrinkage Priors on Variances and Model Parameters

This section describes the priors used in the previously discussed synthesis function. The triple gamma prior can be represented as a conditionally normal distribution, where the component specific variance is itself a compound probability distribution resulting from two gamma distributions. This results in independent normal-gamma-gamma (NGG) priors (Cadonna *et al.*, 2020), both on the standard deviations of the innovations, that is the  $\sqrt{\theta_j}$ 's, and on the means of the initial value  $\beta_j$ , for  $j = 1, \dots, d$ . Note that, in the case of the standard deviations, this can equivalently be seen as a triple gamma prior on the innovation variances  $\theta_j$ , for  $j = 1, \dots, d$ . In the constant parameterizations, I place an NGG prior on the  $\beta_j$  using the centered parameterization:

$$\sqrt{\theta_j}|\xi_j^2 \sim \mathcal{N}(0, \xi_j^2), \quad \xi_j^2|a^\xi, \kappa_j^2 \sim \mathcal{G}(a^\xi, \frac{a^\xi \kappa_j^2}{2}), \quad \kappa_j^2|c^\xi, \kappa_B^2 \sim \mathcal{G}(c^\xi, \frac{c^\xi}{\kappa_B^2}) \quad (\text{A.4})$$

$$\beta_j|\tau_j^2 \sim \mathcal{N}(0, \tau_j^2), \quad \tau_j^2|a^\tau, \lambda_j^2 \sim \mathcal{N}(a^\tau, \frac{a^\tau \lambda_j^2}{2}) \quad \lambda_j^2|c^\tau, \lambda_B^2 \sim \mathcal{N}(c^\tau, \frac{c^\tau}{\lambda_B^2}). \quad (\text{A.5})$$

Letting  $c^\xi$  and  $c^\tau$  go to infinity results in a normal-gamma (NG) prior (Brown and Griffin, 2010) on the  $\sqrt{\theta_j}$ 's and  $\beta_j$ 's. It also has a representation as a conditionally normal distribution, with the component specific variance following a gamma distribution; that is

$$\sqrt{\theta_j}|\xi_j^2 \sim \mathcal{N}(0, \xi_j^2), \quad \xi_j^2|a^\xi, \kappa_B^2 \sim \mathcal{G}(a^\xi, \frac{a^\xi \kappa_B^2}{2}), \quad (\text{A.6})$$

$$\beta_j|\tau_j^2 \sim \mathcal{N}(0, \tau_j^2), \quad \tau_j^2|a^\tau, \lambda_B^2 \sim \mathcal{G}(a^\tau, \frac{a^\tau \lambda_B^2}{2}). \quad (\text{A.7})$$

The parameters  $a^\xi$ ,  $a^\tau$ ,  $c^\xi$ ,  $c^\tau$ ,  $\kappa_B^2$ , and  $\lambda_B^2$  can be learned from the data through appropriate prior distributions. Results from Cadonna *et al.* (2020) motivate the use of different distributions for these parameters under the NGG and NG prior. In the NGG case, the scaled global shrinkage parameters conditionally follow F distributions, depending on their respective pole and tail parameters:

$$\frac{\kappa_B^2}{2}|a^\xi, c^\xi \sim F(2a^\xi, 2c^\xi), \quad \frac{\lambda_B^2}{2}|a^\tau, c^\tau \sim F(2a^\tau, 2c^\tau). \quad (\text{A.8})$$

The scaled tail and pole parameters, in turn, follow beta distributions:

$$2a^\xi \sim \mathcal{B}(\alpha_{a^\xi}, \beta_{a^\xi}), \quad 2c^\xi \sim \mathcal{B}(\alpha_{c^\xi}, \beta_{c^\xi}), \quad (\text{A.9})$$

$$2a^\tau \sim \mathcal{B}(\alpha_{a^\tau}, \beta_{a^\tau}), \quad 2c^\tau \sim \mathcal{B}(\alpha_{c^\tau}, \beta_{c^\tau}). \quad (\text{A.10})$$

These priors are chosen as they imply a uniform prior on a suitably defined model size; see Cadonna *et al.* (2020) for details. In the NG case, the global shrinkage param-

eters follow independent gamma distributions:

$$\kappa_B^2 \sim \mathcal{G}(d_1, d_2), \quad \lambda_B^2 \sim \mathcal{G}(e_1, e_2). \quad (\text{A.11})$$

In order to learn the pole parameters in the NG case, I generalize the approach taken in Bitto and Frühwirth-Schnatter (2019) and place the following gamma distributions as priors:

$$a^\xi \sim \mathcal{G}(\alpha_{a^\xi}, \alpha_{a^\xi} \beta_{a^\xi}), \quad a^\tau \sim \mathcal{G}(\alpha_{a^\tau}, \alpha_{a^\tau} \beta_{a^\tau}), \quad (\text{A.12})$$

which correspond to the exponential priors used in Bitto and Frühwirth-Schnatter (2019) when  $\alpha_{a^\xi} = 1$  and  $\alpha_{a^\tau} = 1$ . The parameters  $\alpha_{a^\xi}$  and  $\alpha_{a^\tau}$  act as degrees of freedom and allow the prior to be bounded away from zero.

In the constant parameter case, I employ a hierarchical prior, where the scale of an inverse gamma prior for  $\sigma^2$  follows a gamma distribution; that is,

$$\sigma^2 | C_0 \sim \mathcal{G}^{-1}(c_0, C_0), \quad C_0 \sim \mathcal{G}(c_0 + g_0, (G_0 + \sigma^{-2})^{-1}), \quad (\text{A.13})$$

with hyperparameters  $c_0$ ,  $g_0$ , and  $G_0$ .

In the case of stochastic volatility, the priors on the parameters  $\sigma_\eta^2$  in Equation A.3 are,

$$\sigma_\eta^2 \sim \mathcal{G}^{-1}(\nu, S_h), \quad h_0 \sim \mathcal{N}(a_0, b_0) \quad (\text{A.14})$$

with hyperparameters  $\nu$ ,  $S_h$ ,  $a_0$  and  $b_0$ .

### MCMC Sampling Algorithm

This next section describes the MCMC Gibbs sampling algorithm with Metropolis-Hastings steps to obtain draws from the posterior distribution of the global-local shrinkage prior synthesis function parameters. This is meant to be an overview of the algorithm; for more details, please refer to Cadonna *et al.* (2020) and Bitto and Frühwirth-Schnatter (2019).

**Algorithm 1: Gibbs Sampling Algorithm**

1. If in TVP specification, sample the latent states  $\tilde{\beta} = (\tilde{\beta}_0, \dots, \tilde{\beta}_T)$  in the non-centered parametrization from a multivariate normal distribution using precision sampling (Chan and Jeliazkov, 2009). Otherwise skip.
2. If in TVP specification, sample jointly  $\beta_1, \dots, \beta_d$ , and  $\sqrt{\theta_1}, \dots, \sqrt{\theta_d}$  in the non-centered parametrization from a multivariate normal distribution. Otherwise, sample  $\beta_1, \dots, \beta_d$ , in the centered parameterization from a multivariate normal distribution.
3. If in TVP specification, perform an ancillarity-sufficiency interweaving step and re-draw each  $\beta_1, \dots, \beta_d$  from a normal distribution and each  $\theta_1, \dots, \theta_d$  from a generalized inverse Gaussian distribution using the MATLAB implementation (Hartkopf, 2022) of Hörmann and Leydold (2014). Otherwise skip.
4. Sample (where required) the prior variances  $\xi_1^2, \dots, \xi_d^2$  and  $\tau_1^2, \dots, \tau_d^2$  and the component specific hyper-parameters. Sample the pole, tail, and global shrinkage parameters. In the NGG case, this is done by employing steps (b)–(f) from Algorithm 1 in Cadonna *et al.* (2020). In the NG case, use steps (d) and (e) from Algorithm 1 in Bitto and Frühwirth-Schnatter (2019).
5. Sample the error variance  $\sigma^2$  from an inverse gamma distribution in the homoscedastic case or, in the SV case, sample the volatility of the volatility  $\sigma_\eta^2$  and the log-volatilities  $h = (h_0, \dots, h_T)$ .

Step 4 presents a fork in the algorithm, as different parameterizations are used in the NGG and NG case, to improve mixing. For details on the exact parameterization used in the NGG case, see Cadonna *et al.* (2020). One key feature of the algorithm is the joint sampling of the time-varying parameters  $\tilde{\beta}_t$ , for  $t = 0, \dots, T$  in step 1 of Algorithm 1. I employ the procedure described in Chan and Jeliazkov (2009) and McCausland *et al.* (2011) from Rue and Held (2005), which exploits the sparse, block tri-diagonal structure of the precision matrix of the full conditional distribution of  $\tilde{\beta} = (\tilde{\beta}_0, \dots, \tilde{\beta}_T)$ , to speed up computations.

Step 3, as described in Bitto and Frühwirth-Schnatter (2019), makes use of the ancillarity-sufficiency interweaving strategy (ASIS) introduced by Yu and Meng (2011). ASIS is well known to improve mixing by sampling certain parameters both in the centered and non-centered parameterization.

### A.2.3 Factor Model Combinations

The second synthesis function considered in this paper is a Bayesian Factor Model similar to that of Lopes and West (2004), and Lopes (2014) provides an overview of Modern Bayesian Factor Analysis. Please refer to those references for detailed discussion on the methods. Here I provide a brief overview of the model and estimation technique.<sup>1</sup>

$$y_t = F_t' \gamma_t + \epsilon_t \quad \gamma_t = \gamma_{t-1} + u_t \quad x_t = \Lambda f_t + \nu_t \quad (\text{A.15})$$

$$\epsilon_t \sim \mathcal{N}(0, \sigma_t^2) \quad u_t \sim \mathcal{N}(0, \theta) \quad \nu_t \sim \mathcal{N}(0, R) \quad (\text{A.16})$$

where  $f_t$  is a  $k \times 1$  vector of factors,  $F_t = (1, f_t)$ ,  $\gamma_t$  is  $k + 1$  vector of coefficients,  $\Lambda$  is a  $J \times k$  vector of loadings, and  $R$  is a diagonal covariance matrix with elements  $\sigma_{\nu_j}^2$ . In order to derive combination weights, I need to identify the factors. This is done by the following restriction  $f_t' f_t = I_J$  and by restricting the first  $k$  elements of the loadings matrix to be positive block lower diagonal. This is a common identification scheme used to fix indeterminacy in the estimation of the factors.

To complete model specifications, I need priors for  $\Lambda$ ,  $R$ ,  $\sigma_t^2$ , and  $\theta$ . The factor loadings have independent priors  $\Lambda_{ij} \sim \mathcal{N}(0, C_0)$  when  $i \neq j$  and  $\Lambda_{ij} \sim \mathcal{N}(0, C_0) 1(\Lambda_{ii} > 0)$  for the upper-diagonal elements of positive loadings  $i = 1, \dots, k$ . Each of the prior variances are independent and modeled as  $\sigma_{\nu_j}^2 \sim \mathcal{IG}(\nu/2, \nu s^2/2)$ , similarly  $\theta \sim \mathcal{IG}(\nu_\theta/2, \nu_\theta s_\theta^2/2)$ . Initial conditions for the  $\gamma_t$  are  $\gamma_0 \sim \mathcal{N}(0, P_0)$ , where  $P_0 \sim \mathcal{IG}(\nu_P, (\nu_P - 1) \times c_P)$ .

With the model specified, the next section provides a sketch of the MCMC routine. Interested readers can refer to Lopes and West (2004).

---

<sup>1</sup>The implementation in the paper includes an intercept. For ease of exposition, it has been omitted in the following section.

**Algorithm 2: Gibbs Sampling Algorithm**

1. Sample  $f_t$  from independent normal distributions for every  $t$ , namely,
 
$$f_t \sim \mathcal{N}((I_k + \Lambda'R^{-1}\Lambda)^{-1}\Lambda'R^{-1}x_t, (I_k + \Lambda'R^{-1}\Lambda)^{-1}).$$
2. Sample  $\Lambda$  for  $i = 1, \dots, k$   $\Lambda_i \sim \mathcal{N}(m_i, C_i)1(\Lambda_{ii} > 0)$  where  $m_i = C_i(C_0^{-1}\mu_01_i + \sigma_{\nu_i}^2 F_i x_i)$  and  $C_i^{-1} = C_0^{-1}I_i + \sigma_{\nu_i}^2 F_i' F_i$ .
3. Sample  $\Lambda$  for  $i = k+1, \dots, J$   $\Lambda_i \sim \mathcal{N}(m_i, C_i)1(\Lambda_{ii} > 0)$  where  $m_i = C_i(C_0^{-1}\mu_01_k + \sigma_{\nu_i}^2 F' x_i)$  and  $C_i^{-1} = C_0^{-1}I_k + \sigma_{\nu_i}^2 F' F$ .
4. Sample  $\sigma_{\nu_i}^2 \sim \mathcal{IG}((\nu + T)/2, (\nu s_2 + d_i)/2)$  where  $d_i = (x_i - F\Lambda)'(x_i - F\Lambda)$ .
5. If in TVP specification, sample the latent states  $\gamma_1, \dots, \gamma_d$ , jointly from a multivariate normal distribution using the precision sampler of Chan and Jeliazkov (2009). Otherwise, sample  $\gamma = (\gamma_0, \dots, \gamma_T)$  from a multivariate normal distribution.
6. Sample the error variance  $\sigma^2$  from an inverse gamma distribution in the homoscedastic case or, in the SV case, sample the volatility of the volatility  $\sigma^2$  and the log-volatilities  $h = (h_0, \dots, h_T)$ .

**A.2.4 Sampling the Agent States**

After estimating the synthesis function parameters, the next step in BPS is to draw  $x_{1:t}$  from  $p(x_{1:t}|\Phi_{1:t}, y_{1:t}, \mathcal{H}_{1:t})$  where  $\Phi$  is the model parameters,  $y_t$  is the target variable, and  $\mathcal{H}_{1:t}$  is the set of agent densities. As shown in McAlinn and West (2019), the  $x_t$ , draws from agent densities, are conditionally independent over  $t$  with time  $t$  conditionals:

$$p(x_t|\Phi_t, y_t, \mathcal{H}_t) \propto N(y_t|X_t'\beta_t, \epsilon_t) \prod_{j=1:J} h_{tj}(x_{tj}) \quad \text{with} \quad X_t = (1, x_{t1}, \dots, x_{tJ})' \quad (\text{A.17})$$

If the agents provide normal forecast densities, then A.17 yields a multivariate normal distribution for  $x_t$ . The posterior distribution for each  $x_t$  is:

$$p(x_t|\Phi_t, y_t, \mathcal{H}_t) = \mathcal{N}(h_t + b_t c_t, H_t - b_t b_t' g_t) \quad (\text{A.18})$$

## Appendix A. Chapter 2 Appendix

where  $c_t = y_t - \beta_{t0} - h'_t \beta_{t,1:J}$ ,  $g_t = \sigma_t^2 + \beta'_{t,1:J} H_t \beta_{t,1:J}$ , and  $b_t = H_t \beta_{t,1:J} / g_t$ . Unfortunately, the applications in this paper do not have analytical forms; instead, histograms represent the agent densities. With no analytical form, I use a Block Metropolis-Hastings step with A.18 as a proposal distribution. Since the number of agent densities can be large, I break the MH step into blocks of five experts that are sampled at a time.

There are a few details for Bayesian Factor Model combinations that warrant explanation. First, the model has to be re-parameterized in terms of the  $x_t$  so that I can use the proposal distribution from A.18 in the MH step. The model is straightforward to re-parameterize with the following steps:

$$y_t = x'_t \gamma_t + \epsilon_t \quad x_t = \Lambda f_t + \nu_t \quad (\text{A.19})$$

$$f_t = (\Lambda' \Lambda)^{-1} \Lambda' x_t - (\Lambda' \Lambda)^{-1} \Lambda' \nu_t \quad \text{where,} \quad \Omega = (\Lambda' \Lambda)^{-1} \Lambda' \quad (\text{A.20})$$

$$y_t = x'_t \Omega' \gamma_t - \nu'_t \Omega' \gamma_t + \epsilon_t \quad \rightarrow \quad y_t = x'_t \gamma_t^* + \epsilon_t^* \quad (\text{A.21})$$

$$\text{where,} \quad \epsilon_t^* = -\nu'_t \Omega' \gamma_t + \epsilon_t \quad \text{and} \quad \gamma_t^* = \Omega' \gamma_t \quad (\text{A.22})$$

Now that the model has been re-parameterized, I can use the equation A.18 in the MH step by substituting in  $\beta_t = \gamma_t^*$ , and error variance  $\epsilon_t^* \sim \mathcal{N}(0, \gamma_t^* \Omega R \Omega' \gamma_t + \sigma_t^2)$ .

The second issue that the data ( $x_t$ ) used to estimate Bayesian Factor Models is standardized to be mean 0 and variance 1. Since the agents provide forecast distributions, I calculate mean and variance used to standardized draws from the agent densities using the marginal density of each expert over all  $T$  ( $h(x)_{1:T}$ ). Each  $x_t$  draw is standardized during each MCMC iteration.

### A.3 Calibration Appendix

This section assesses the calibration of the BPS predictions. Calibration (also referred to as absolute accuracy) is achieved when a predictive density properly characterizes the probability of the events that it is predicting. For example, events predicted to occur with a 20 percent probability should be observed in the data roughly 20 percent of the time. More formally, calibration refers to the statistical consistency between the predictive distributions and the observations of the data they are predicting (Gneiting and Raftery, 2007). I assess calibration with a test based off of the probability integral transforms (PITs) (Diebold *et al.*, 1998) as proposed in Knüppel (2015). In general, I find little evidence to suggest that the predictions from any of the synthesis functions are not calibrated.

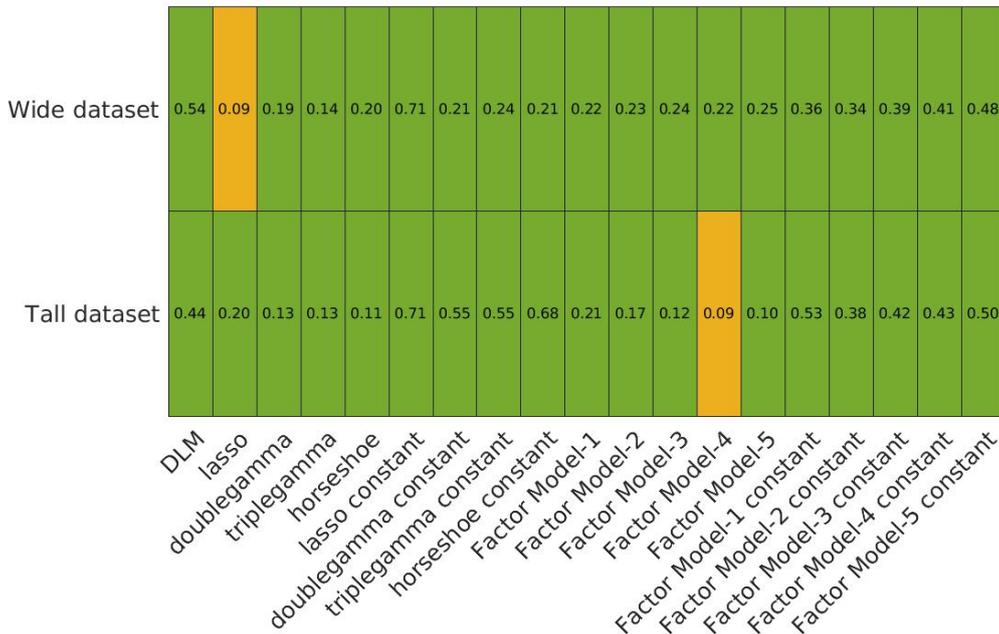
Figures A.1 and A.2 show results from the nowcasting application and the SPF forecasting application. For the most part in the nowcasting applications, the factor model combinations show little evidence of being uncalibrated. However, the shrinkage approaches have slightly different results. The LASSO synthesis function does not appear to produce calibrated predictions, and the test rejects calibration for the time-varying double gamma specification at most horizons. In contrast, the constant parameter specifications produce calibrated predictions at most horizons, the exception being the shortest horizons where calibration is rejected at the 10 percent level. The SPF application has more straightforward results—there is little evidence to suggest the BPS predictions are uncalibrated from any synthesis function. In only two cases is the null hypothesis rejected at the 10 percent level.

Figure A.1: Knüpple Test for Probabilistic Calibration: Nowcasting Application

24 weeks until NA	0.15	0.06	0.12	0.53	0.43	0.09	0.72	0.91	0.81	0.42	0.47	0.44	0.47	0.48	0.96	0.79	0.73	0.65	0.58
22 weeks until NA	0.22	0.06	0.11	0.45	0.45	0.10	0.79	0.96	0.86	0.46	0.44	0.47	0.47	0.48	0.94	0.83	0.77	0.67	0.59
20 weeks until NA	0.28	0.04	0.08	0.38	0.26	0.07	0.60	0.67	0.66	0.53	0.54	0.54	0.57	0.57	0.90	0.68	0.68	0.52	0.49
18 weeks until NA	0.26	0.04	0.07	0.29	0.31	0.07	0.73	0.85	0.70	0.57	0.57	0.59	0.60	0.61	0.90	0.71	0.61	0.51	0.46
16 weeks until NA	0.17	0.08	0.07	0.47	0.33	0.12	0.81	0.96	0.92	0.86	0.87	0.86	0.84	0.86	0.70	0.68	0.61	0.56	0.55
14 weeks until NA	0.17	0.07	0.10	0.58	0.49	0.11	0.82	0.81	0.79	0.92	0.88	0.85	0.86	0.87	0.69	0.62	0.57	0.57	0.53
12 weeks until NA	0.15	0.04	0.05	0.51	0.33	0.08	0.78	0.80	0.81	0.89	0.87	0.90	0.89	0.87	0.76	0.73	0.68	0.60	0.56
10 weeks until NA	0.14	0.03	0.06	0.43	0.35	0.08	0.80	0.84	0.81	0.92	0.88	0.93	0.92	0.91	0.75	0.69	0.68	0.63	0.61
8 weeks until NA	0.45	0.01	0.01	0.05	0.03	0.05	0.13	0.12	0.10	0.63	0.68	0.72	0.71	0.69	0.67	0.54	0.70	0.57	0.49
6 weeks until NA	0.32	0.01	0.00	0.02	0.02	0.04	0.11	0.08	0.08	0.64	0.70	0.73	0.70	0.70	0.64	0.55	0.61	0.48	0.28
4 weeks until NA	0.10	0.06	0.06	0.33	0.24	0.07	0.03	0.06	0.06	0.55	0.50	0.43	0.34	0.33	0.18	0.07	0.05	0.03	0.02
2 weeks until NA	0.12	0.07	0.06	0.42	0.17	0.08	0.03	0.07	0.05	0.56	0.58	0.45	0.40	0.34	0.14	0.08	0.09	0.08	0.12
	DLM	lasso	doublegamma	triplegamma	horseshoe	lasso constant	doublegamma constant	triplegamma constant	horseshoe constant	Factor Model-1	Factor Model-2	Factor Model-3	Factor Model-4	Factor Model-5	Factor Model-1 constant	Factor Model-2 constant	Factor Model-3 constant	Factor Model-4 constant	Factor Model-5 constant

Notes: Results from the Knüpple test for probabilistic calibration. Null hypothesis is for calibration and values in the table correspond to p-values. Red shading corresponds to rejection of calibration at 5 percent level and yellow at 10 percent level.

Figure A.2: Knüpple Test for Probabilistic Calibration: Survey Forecast Application



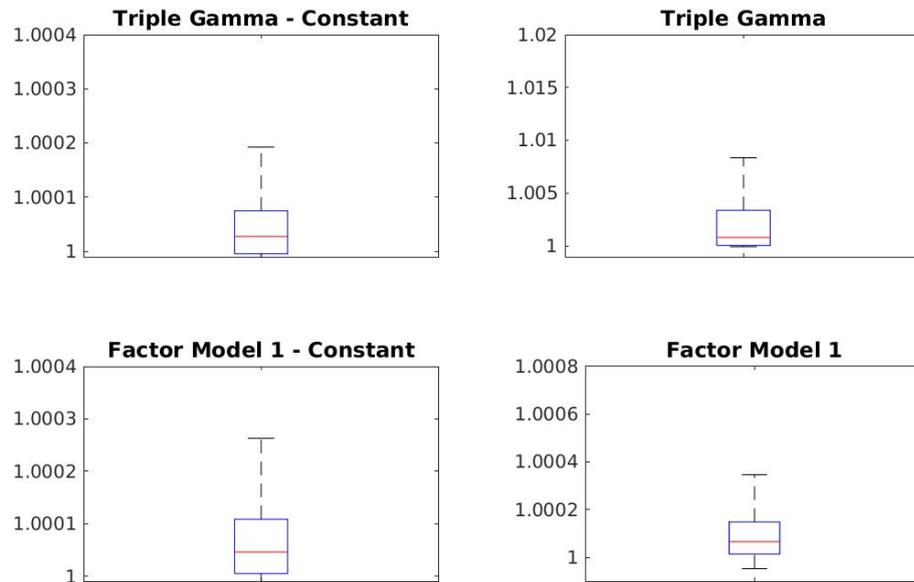
Notes: Results from the Knüpple test for probabilistic calibration. Null hypothesis is for calibration and values in the table correspond to p-values. Red shading corresponds to rejection of calibration at 5 percent level and yellow at 10 percent level.

## A.4 MCMC Convergence Appendix

In this section, I assess the convergence of the MCMC algorithms. This is done using the Gelman-Rubin diagnostic (Rubin *et al.*, 2015) and implemented through the MATLAB MCMC Diagnostics Toolbox (Vehtari and Särkkä, 2014). The Gelman-Rubin diagnostic compares within-chain variance to across-chain variance to estimate a potential scale reduction factor (R), which can be used to assess convergence of the MCMC chain. As a rule of thumb, values below 1.1 suggest convergence. The diagnostic is performed with five chains and on four specifications of BPS using the SPF “tall” dataset. I focus on the constant and time-varying versions of the triple gamma and one factor synthesis functions. This is because the other shrinkage priors are special cases of the triple gamma prior and reduce to simpler versions of the sampler. Results for other shrinkage are available upon request. Since the number of parameters, state variables, and hyper-

parameters sampled can number in the thousands, I report box plots of the potential scale reduction factor in Figure A.3. These results show reasonable convergence of the MCMC algorithms.

Figure A.3: Box Plots of Potential Scale Reduction Factors



Notes: The above shows the potential scale reduction factors (R) from the Gelman-Rubin diagnostic with 5 chains on the SPF 'Tall' dataset. Values less than 1.1 provide evidence that the MCMC has converged.

## Appendix B

### Chapter 3 Appendix

## B.1 Technical Appendix: Bayesian Inference

Before we start discussing the modeling choices, and the prior and the posterior sampler in detail, we introduce a bit of additional notation to simplify the exposition.

We can rewrite Eq. (3.2) as a standard TVP regression:

$$y_t = c_t + \boldsymbol{\gamma}' \mathbf{x}_{t|t-h} + \boldsymbol{\beta}'_t \mathbf{x}_{t|t-h} + \sigma_t \nu_t, \quad \nu_t \sim \mathcal{N}(0, 1). \quad (\text{B.1})$$

The priors on  $\boldsymbol{\gamma}$  and  $\boldsymbol{\beta}_t$  can be written in the form of multivariate Gaussian distributions:

$$\begin{aligned} \boldsymbol{\gamma} &\sim \mathcal{N}(\boldsymbol{\mu}^\gamma(\mathbf{Z}^\gamma), \mathbf{V}^\gamma), \\ \boldsymbol{\beta}_t &\sim \mathcal{N}(\boldsymbol{\mu}^\beta(\mathbf{z}_{t|t-h}^\beta), \mathbf{V}^\beta). \end{aligned}$$

Here,  $\boldsymbol{\mu}^\gamma(\mathbf{Z}^\gamma)$  and  $\boldsymbol{\mu}^\beta(\mathbf{Z}_{t|t-h}^\beta)$  are both prior mean functions of dimension  $J$  and  $\mathbf{V}^n = \text{diag}(\tau_1^n, \dots, \tau_J^n)$  for  $n \in \{\gamma, \beta\}$ . The weight modifiers are stored, respectively, in a  $(J \times K_\gamma)$  matrix  $\mathbf{Z}^\gamma$  with typical row  $\mathbf{z}_j^\gamma$  and in a  $(JT \times K_\beta)$  matrix  $\mathbf{Z}^\beta = (\mathbf{z}_1^\beta, \dots, \mathbf{z}_T^\beta)'$ , with  $\mathbf{z}_t^\beta = (\mathbf{z}_{1t|t-h}^\beta, \dots, \mathbf{z}_{Jt|t-h}^\beta)$ ,  $t = 1, \dots, T$ .

Notice that the prior on stacked  $\boldsymbol{\beta} = (\boldsymbol{\beta}'_1, \dots, \boldsymbol{\beta}'_T)$  can be written as a  $JT$ -dimensional Gaussian distribution:

$$\boldsymbol{\beta} \sim \mathcal{N}(\boldsymbol{\mu}^\beta(\mathbf{Z}^\beta), \mathbf{I}_T \otimes \mathbf{V}^\beta),$$

where the prior mean function  $\boldsymbol{\mu}^\beta(\mathbf{Z}^\beta)$  is now also of dimension  $JT$ .

### B.1.1 Additional Details about Our Modeling Choices and Priors

In this sub-section we provide additional details about our hierarchical prior setup used for  $\boldsymbol{\gamma}$  and the time-varying part,  $\boldsymbol{\beta}_t$ , of the weights. In both cases, the prior mean functions,  $\boldsymbol{\mu}^\gamma(\mathbf{Z}^\gamma)$  and  $\boldsymbol{\mu}^\beta(\mathbf{z}_t^\beta)$ , are approximated by tree functions (see Eq. 3.6), while the prior variances – which define the degree of shrinkage toward these prior means – are modeled with a horseshoe (HS, Carvalho *et al.*, 2010) prior. In addition, we sketch

the law of motion and modeling choices for the time-varying intercept and time-varying variances in Eq. (B.1), both of which capture the idea of model incompleteness.

**Tree functions to approximate the prior mean.** We closely follow here the suggestions of the Bayesian additive regression tree (BART) literature (Chipman *et al.*, 1998, 2010) and use a similar prior setup for our tree structures  $\mathcal{T}_s^n$  and terminal node parameters  $\phi_s^n$  for  $n \in \{\beta, \gamma\}$ . To generate the tree function, Chipman *et al.* (1998) and Chipman *et al.* (2010) suggest using a stochastic process of the following form:

1. **Prior on the tree structure  $\mathcal{T}_s^n$ .** Impose a decreasing probability of growing more complex trees and that a terminal node is non-terminal. This probability is assumed to be

$$\frac{c_0}{(1 + \vartheta)^{c_1}},$$

for a particular terminal node at depth  $\vartheta$ , with the hyperparameters  $c_0 = 0.95$  and  $c_1 = 2$  being two values that have been shown to be reasonable choices in much of the literature using Bayesian (additive) tree models. Chipman *et al.* (2010) show that this choice works well even for single-tree models. Moreover, for each splitting rule at each node, Chipman *et al.* (2010) propose a prior that is agnostic about the choice of the specific splitting variable and propose a natural default choice, which is to use a uniform prior on the splitting variables, treating each variable as equally likely to be used in a splitting rule.

2. **Prior on the terminal node parameters  $\phi_s^n$ .** We use a Gaussian prior for the terminal node parameters. For a typical element in  $\phi_s^n$ , that is

$$\phi_{j,s}^n \sim \mathcal{N}(0, c_2/S),$$

where  $c_2$  refers to a shrinkage parameter and  $S$  to the number of trees. It is worth noting that – to avoid overfitting – the prior variances for these terminal parameters

are scaled down by the number of trees and become tighter, so that each individual tree explains only a tiny fraction within the additive sum-of-tree function.

**Shrinkage toward the prior mean through the horseshoe prior.** The horseshoe prior amounts to setting the scaling parameters as follows:

$$\tau_j^n = \lambda^n \psi_j^n, \quad \lambda^n \sim \mathcal{C}^+(0, 1), \quad \psi_j^n \sim \mathcal{C}^+(0, 1), \quad \text{for } n \in \{\gamma, \beta\},$$

with  $\mathcal{C}^+(0, 1)$  denoting the half-Cauchy distribution. The key feature of this prior is that  $\lambda^n$  serves as a global shrinkage parameter that pulls all weights toward the prior mean, whereas  $\psi_j^n$  allows for agent-specific deviations from this common pattern. Another representation of this prior, which simplifies posterior sampling enormously, is based on introducing inverse Gamma distributed auxiliary variables (see Makalic and Schmidt, 2015):

$$\begin{aligned} \lambda^n | \zeta^n &\sim \mathcal{G}^{-1}(1/2, 1/\varphi^n), & \varphi^n &\sim \mathcal{G}^{-1}(1/2, 1), \\ \psi_j^n | \varpi_j^n &\sim \mathcal{G}^{-1}(1/2, 1/\varpi_j^n), & \varpi_j^n &\sim \mathcal{G}^{-1}(1/2, 1). \end{aligned}$$

This representation is convenient since – when combined with the likelihood – it gives rise to a simple Gibbs sampling step that involves only inverse Gamma full conditionals (see Sub-section B.1.2 below).

**Controlling for model incompleteness.** A time-varying intercept  $c_t$  and time-varying variances  $\sigma_t^2$  both control for model incompleteness. The fact that both are potentially time-varying gives additional flexibility in the degree of model set incompleteness (as outlined in Sub-section 3.2.1). The time-varying intercept follows a random walk (RW) law of motion with the state equation given by

$$c_t = c_{t-1} + \eta_{c,t}, \quad \eta_{c,t} \sim \mathcal{N}(0, \sigma_c^2),$$

where  $\sigma_c^2$  denotes the state innovation variance. To discipline  $c_t$ , we use a relatively tight Gamma prior on  $\sigma_c^2$  and strongly push the state innovation variance toward a small positive value close to zero.

The error variances  $\sigma_t^2$  in Eq. (3.1) can be time-varying or constant. The time-varying case is given by

$$\varsigma_t = \mu_\varsigma + \rho_\varsigma(\varsigma_{t-1} - \mu_\varsigma) + \eta_{\varsigma,t}, \quad \eta_{\varsigma,t} \sim \mathcal{N}(0, \sigma_\varsigma^2), \quad (\text{B.2})$$

with  $\mu_\varsigma$  denoting the unconditional mean,  $\rho_\varsigma$  the persistence parameter, and  $\sigma_\varsigma^2$  the state innovation variance of the log-volatility process. For SV we follow Kastner and Frühwirth-Schnatter (2014) and assume a Gaussian prior on  $\mu_\varsigma \sim \mathcal{N}(0, 10^2)$ , a (transformed) Beta prior on  $(\rho_\varsigma + 1)/2 \sim \mathcal{B}(5, 1.5)$ , and a Gamma prior on  $\sigma_\varsigma^2 \sim \mathcal{G}(0.5, 0.5)$ . Moreover, for the case of homoskedastic errors, we assume an inverse Gamma prior on  $\sigma^2 \sim i\mathcal{G}(0.01, 0.01)$ .

### B.1.2 Posterior Simulation

The prior discussed in the previous section can be combined with the likelihood to derive the full posterior over all unknown quantities in our model. Since this joint density is untractable, we use Markov chain Monte Carlo (MCMC) methods to carry out posterior simulation. In what follows, we let  $\bullet$  be generic notation that indicates that we condition on all other parameters/states of the model.

We start the discussion of our posterior sampler by first describing how we estimate the latent quantities that enter the synthesis function. This includes the static and dynamic weights, the error variances, the latent trend components, and the agent-specific forecasts.

**Sampling from  $\{p(c_t|\bullet)\}_{t=1}^T$ .** We sample the full history of the random walk intercept term conditional on all other unknowns in the model using a forward-filtering backward-sampling (FFBS) step (Carter and Kohn, 1994). This is achieved by noting

that

$$\underbrace{y_t - \gamma' \mathbf{x}_{t|t-h} - \beta_t' \mathbf{x}_{t|t-h}}_{y_t^{\gamma, \beta}} = c_t + \sigma_t \nu_t \quad (\text{B.3})$$

is a standard unobserved components model with heteroskedastic shocks.

**Sampling from  $p(\gamma|\bullet)$ .** The time-invariant weights are sampled from a  $J$ -dimensional Gaussian full conditional posterior distribution,

$$\gamma|\bullet \sim \mathcal{N}(\bar{\gamma}, \bar{\mathbf{V}}^\gamma), \quad (\text{B.4})$$

with moments given by

$$\begin{aligned} \bar{\mathbf{V}}^\gamma &= (\mathbf{X}' \boldsymbol{\Sigma}^{-1} \mathbf{X} + (\underline{\mathbf{V}}^\gamma)^{-1})^{-1}, \\ \bar{\gamma} &= \bar{\mathbf{V}}^\gamma \left( \mathbf{X}' \boldsymbol{\Sigma}^{-1} \mathbf{y}^{\beta, c} + (\underline{\mathbf{V}}^\gamma)^{-1} \mu^\gamma(\mathbf{Z}^\gamma) \right), \end{aligned}$$

where  $\mathbf{X}$  is a  $T \times J$  matrix with  $t^{\text{th}}$  row  $\mathbf{x}'_{t|t-h}$ ,  $\boldsymbol{\Sigma} = \text{diag}(\sigma_1^2, \dots, \sigma_T^2)$  and  $\mathbf{y}^{\beta, c}$  is a  $T$ -dimensional vector with typical element  $y_t - \beta_t' \mathbf{x}_{t|t-h} - c_t$ .

**Sampling  $p(\beta|\bullet)$ .** The dynamic regression coefficients are simulated by writing the model in static form. The static form of the model reads:

$$\mathbf{y}^{\gamma, c} = \mathbf{W} \boldsymbol{\beta} + \boldsymbol{\nu}, \quad \boldsymbol{\nu} \sim \mathcal{N}(\mathbf{0}_T, \boldsymbol{\Sigma}), \quad (\text{B.5})$$

where  $\mathbf{y}^{\gamma, c}$  is  $T \times 1$  and has typical element  $y_t - \gamma' \mathbf{x}_{t|t-h} - c_t$  and  $\mathbf{W}$  is a  $T \times TJ$ -dimensional block diagonal matrix with  $\mathbf{W} = \text{bdiag}(\mathbf{x}'_{1|1-h}, \dots, \mathbf{x}'_{T|T-h})$ .<sup>1</sup> Under this static representation, the posterior of  $\boldsymbol{\beta}$  takes a standard form and is multivariate Gaussian:

$$\boldsymbol{\beta}|\bullet \sim \mathcal{N}(\bar{\boldsymbol{\beta}}, \bar{\mathbf{V}}^\beta), \quad (\text{B.6})$$

---

<sup>1</sup>Observations  $-h, \dots, 0$  refer to a part of the sample that we use to initialize our models.

with posterior covariance matrix and mean vector given by, respectively,

$$\begin{aligned}\bar{\mathbf{V}}^\beta &= \left( \mathbf{W}' \boldsymbol{\Sigma}^{-1} \mathbf{W} + (\mathbf{I}_T \otimes \underline{\mathbf{V}}^\beta)^{-1} \right)^{-1}, \\ \bar{\boldsymbol{\beta}} &= \bar{\mathbf{V}}^\beta \left( \mathbf{W}' \boldsymbol{\Sigma}^{-1} \mathbf{y}^{\gamma, c} + (\mathbf{I}_T \otimes \underline{\mathbf{V}}^\beta)^{-1} \boldsymbol{\mu}^\beta(\mathbf{Z}^\beta) \right).\end{aligned}$$

This distribution is high dimensional even for moderate values of  $J$ , and we thus use the efficient sampler outlined in Hauzenberger *et al.* (2022).

**Sampling from  $p(\sigma_1^2, \dots, \sigma_T^2 | \bullet)$ .** We sample the log volatilities and associated state equation parameters using the algorithm outlined in Kastner and Frühwirth-Schnatter (2014). This step is implemented in the R package `stochvol`.

**Sampling from  $\{p(\mathbf{x}_{t|t-h} | \bullet)\}_{t=1}^T$ .** We draw from  $\{p(\mathbf{x}_{t|t-h} | \bullet)\}_{t=1}^T$  on a  $t$ -by- $t$  basis. The time  $t$  full conditional posterior of  $\mathbf{x}_t$  is given by

$$p(\mathbf{x}_{t|t-h} | \bullet) \propto \mathcal{N}(y_t | c_t + \boldsymbol{\gamma}' \mathbf{x}_{t|t-h} + \boldsymbol{\beta}'_t \mathbf{x}_{t|t-h}, \sigma_t^2) \prod_{j=1:J} \pi_{jt}(x_{jt|t-h}),$$

which, unless the agent densities  $\pi_{jt}(x_{jt|t-h})$  are Gaussian, takes no well-known form. In our applications, the agent densities do not have analytical representations. For example, the ECB-SPF elicits histograms from survey respondents and in the US application the available forecasts are predictive draws based on model-specific Gibbs samplers. Accordingly, we sample  $\mathbf{x}_t$  using an adaptive Metropolis Hastings step (see, e.g., Roberts and Rosenthal, 2009). This step proposes  $\mathbf{x}_{t|t-h}^*$  from a mixture of Gaussian distributions:

$$\mathbf{x}_{t|t-h}^* \sim (1 - \kappa) \mathcal{N}(\mathbf{x}_{t|t-h}, (2.38)^2 \hat{\mathbf{Q}}_{tm}/J) + \kappa \mathcal{N}(\mathbf{x}_{t|t-h}, (0.1)^2 \mathbf{I}_J/J), \quad (\text{B.7})$$

where  $\kappa = 0.05$  is a small constant and  $\hat{\mathbf{Q}}_{tm}/J$  is the empirical covariance matrix of the target distribution based on the first  $m$  draws. Since this algorithm learns the proposal, it can quickly adjust to cases where the agent densities are non-Gaussian, feature multiple modes, or are heavy tailed.

Next, we discuss the steps involved in sampling the parameters of the priors on the weights.

**Sampling from  $p(\mathcal{T}_1^n, \dots, \mathcal{T}_S^n, \phi_1^n, \dots, \phi_S^n | \bullet)$  for  $n \in \{\gamma, \beta\}$ .** We sample the tree structures and the terminal node parameters using the algorithm proposed in Chipman *et al.* (2010). This algorithm is applicable since, conditional on  $\gamma$  and  $\beta$ , the corresponding priors can be interpreted as regression models. For instance, in the case of  $\beta$ , notice that

$$\beta_{jt} = \sum_{s=1}^S g(\mathbf{z}_{jt}^\beta | \mathcal{T}_s^\beta, \phi_s^\beta) + \tau_j^\beta \nu_{jt}, \quad \nu_{jt} \sim \mathcal{N}(0, 1), \quad (\text{B.8})$$

which, in stacked form, can be written as

$$\beta = \sum_{s=1}^S g(\mathbf{Z}^\beta | \mathcal{T}_s^\beta, \phi_s^\beta) + \mathbf{r}, \quad \mathbf{r} \sim \mathcal{N}(\mathbf{0}_{TJ}, \mathbf{I}_T \otimes \mathbf{V}^\beta). \quad (\text{B.9})$$

Equation B.9 is a standard BART regression with latent responses and heteroskedastic errors. For  $\gamma$ , a similar regression representation can be derived.

**Sampling from  $p(\tau_1^n, \dots, \tau_J^n | \bullet)$  for  $n \in \{\gamma, \beta\}$ .** The scaling parameters are obtained using the algorithm described in Makalic and Schmidt (2015). This algorithm involves only inverse Gamma distributions and, for brevity, we do not discuss them in detail here.

These steps form our MCMC algorithm. In all our empirical work, we iteratively sample from the different full conditionals to obtain draws from the joint posterior of the coefficients and the latent states. Based on these draws, we back out the predictive distribution as described in the main text through Monte Carlo integration. That is, after obtaining a draw from the posterior, we use this draw to forecast  $y_{t^*}$ . This is done for every draw, leading to a posterior distribution over future values  $y_{t^*}$ . In all our empirical work, we repeat this 12,500 times and discard the initial 2,500 draws as a burn-in. We subsequently keep every second draw, yielding a total of 5,000 draws from the joint posterior distribution.

### B.1.3 ADL Model Estimation

Our US inflation forecasting exercise involves Bayesian estimation of ADL models involving different explanatory variables. Table B.1 provides an overview of the 27 variables each used as an exogenous predictor in Eq. (3.12) and also highlights (in bold typeface) the target variable.

Table B.1: List of variables used for the autoregressive distributed lag (ADL) specifications.

Mnemonic	Description	Transformation
GDP1	Real Gross Domestic Product	$100 \times \Delta \log$
PCECC96	Real Personal Consumption Expenditures	$100 \times \Delta \log$
FPIx	Real Private Fixed Investment	$100 \times \Delta \log$
GCEC1	Real Government Consumption Expenditures and Gross Investment	$100 \times \Delta \log$
INDPRO	Total index Industrial Production Index	$100 \times \Delta \log$
CUMFNS	Capacity Utilization: Manufacturing (SIC)	none
PAYEMS	Emp:Nonfarm All Employees: Total nonfarm	$100 \times \Delta \log$
CE16OV	Civilian Employment	$100 \times \Delta \log$
UNRATE	Civilian Unemployment Rate	$\Delta$
AWHMAN	Average Weekly Hours of Production and Nonsupervisory Employees: Manufacturing Hours	none
CES0600000007	Average Weekly Hours of Production and Nonsupervisory Employees: Goods-Producing	$\Delta$
CLAIMSx	Initial Claims	$100 \times \Delta \log$
GDPCTPI	Gross Domestic Product: Chain-type Price Index	$100 \times \Delta \log$
<b>CPIAUCSL</b>	<b>Consumer Price Index for All Urban Consumers</b>	$100 \times \Delta \log$
PPIACO	Producer Price Index for All Commodities	$100 \times \Delta \log$
WPSID61	Producer Price Index by Commodity Intermediate Materials: Supplies & Components	$100 \times \Delta \log$
WPSID62	Producer Price Index: Crude Materials for Further Processing	$100 \times \Delta \log$
COMPRNFB	Nonfarm Business Sector: Real Compensation Per Hour	$100 \times \Delta \log$
ULCNFB	Nonfarm Business Sector: Unit Labor Cost	$100 \times \Delta \log$
CES0600000008	Average Hourly Earnings of Production and Nonsupervisory Employees	$100 \times \Delta \log$
FEDFUNDS	Effective Federal Funds Rate	$\Delta$
BAA10YM	Moody's Seasoned BAA Corporate Bond Yield Relative to Yield on 10-Year Treasury	none
GS10TB3Mx	10-Year Treasury Constant Maturity Minus 3-Month Treasury Bill, secondary market	none
CPF3MTB3Mx	3-Month Commercial Paper Minus 3-Month Treasury Bill, secondary market	none
M2REAL	Real M2 Money Stock	$100 \times \Delta \log$
BUSLOANSx	Real Commercial and Industrial Loans, All Commercial Banks	$100 \times \Delta \log$
CONSUMERx	Real Consumer Loans at All Commercial Banks	$100 \times \Delta \log$
S.P.500	S&Ps Common Stock Price Index	$100 \times \Delta \log$

**Notes:** The variable in bold refers to the target inflation series.

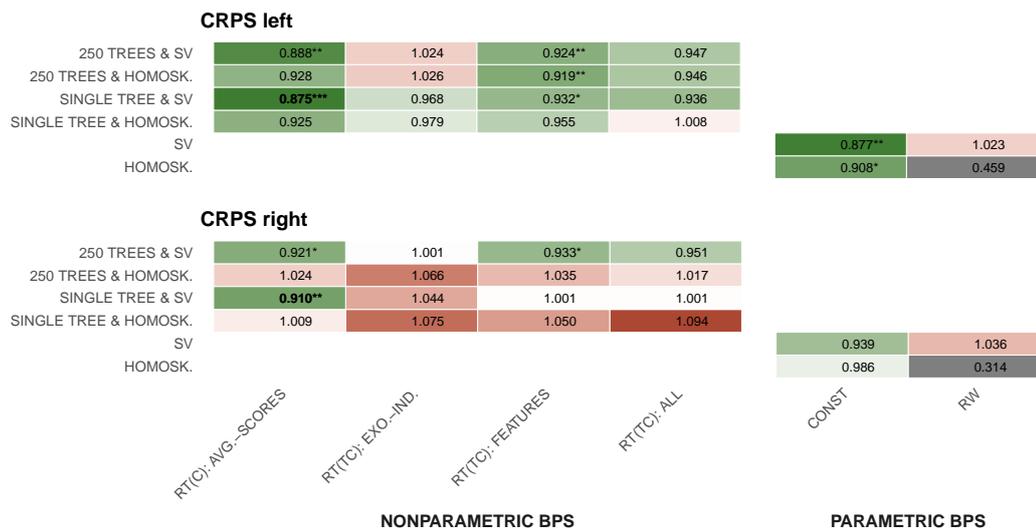
To estimate these ADL specifications, we use standard Bayesian non-conjugate regression techniques with posteriors of standard form. The non-conjugate priors are weakly informative. We center both  $\rho_\pi$  and  $\alpha_\pi$  in Eq. (3.12) on a prior mean of zero and assume a prior variance of 100. For the case of homoskedastic errors, we assume an inverse Gamma prior on  $\sigma_{\pi,t+h}^2 := \sigma_\pi^2 \sim i\mathcal{G}(0.01, 0.01)$ , while for SV we essentially use the setup sketched in Appendix B.1.1 (see Eq. B.2). However, non-conjugacy (and SV in particular) leads to predictive densities for which there is no closed-form solution. Therefore we use MCMC methods and predictive simulation.

## B.2 Empirical Appendix: Additional Results

This empirical appendix contains supplementary results as referenced in the main paper. It is structured as follows. Section B.2.1 reports left and right tail CRPSs. Section B.2.2 presents the combination weights. Section B.2.3 presents the cumulative CRPS statistics. Section B.2.4 presents the fluctuation tests. Section B.2.5 presents the PITs tests. Section B.2.6 presents results showing how we can draw out the degree of shrinkage implied by BPS-RT. Section B.2.7 plots the predictive densities in both applications and examines their skewness.

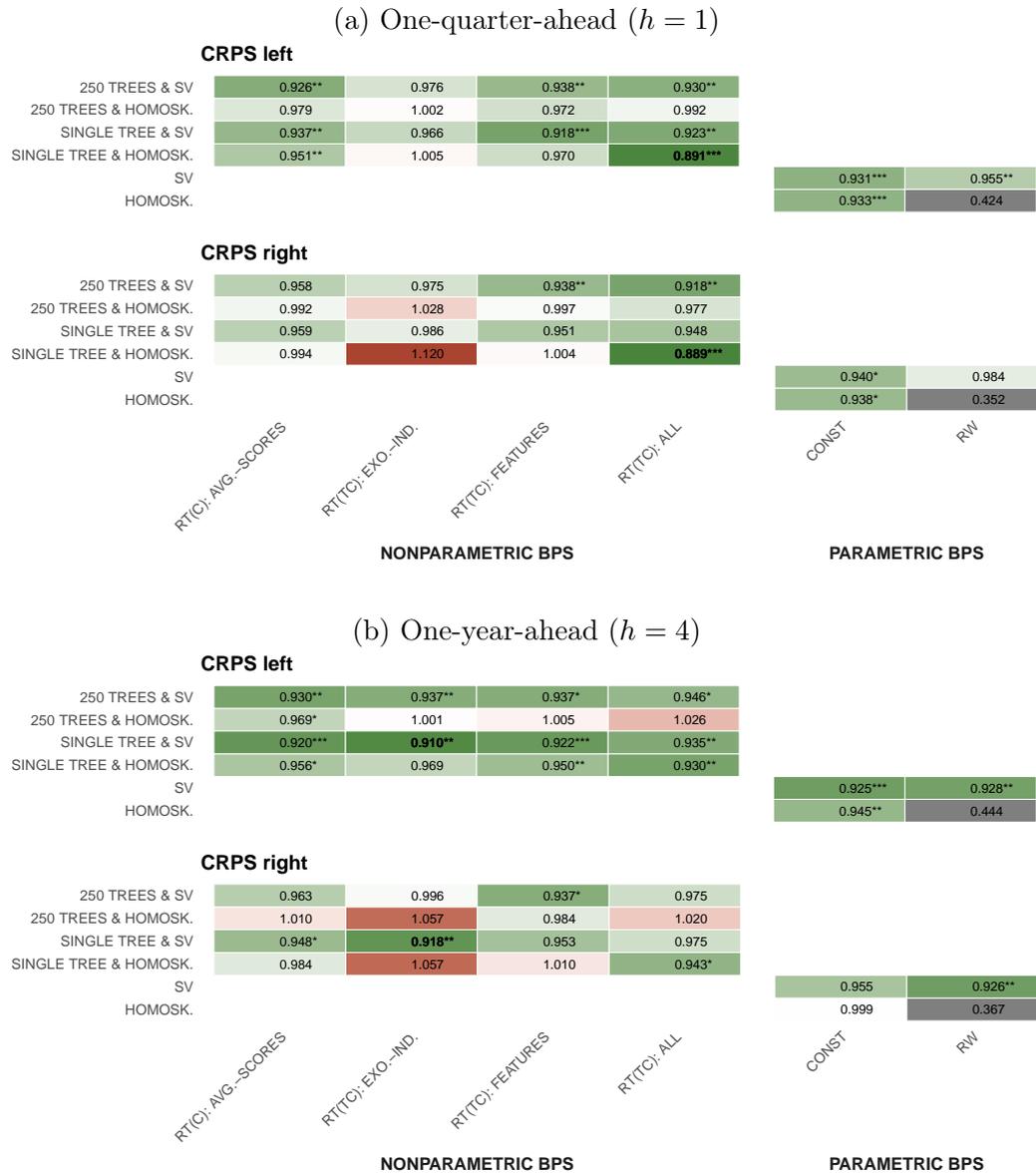
### B.2.1 Tail forecast accuracy

Figure B.1: Relative tail forecast accuracy: EA GDP growth.



**Notes:** This figure shows two variants of quantile-weighted CRPS ratios, one focusing on the left tail and the other on the right tail. The gray-shaded entries give the actual scores of our benchmark (BPS-RW with homoskedastic error variances). Green-shaded entries refer to models that outperform the benchmark (with the forecast metric ratios below one), while red-shaded entries denote models that are outperformed by the benchmark (with the forecast metric ratios greater than one). The best-performing model specification by forecast metric is given in bold. Asterisks indicate statistical significance of the Diebold and Mariano (1995) test, which assumes equal forecast performance for each model relative to the benchmark, at the 1 (\*\*\*) , 5 (\*\*), and 10 (\*) percent significance levels.

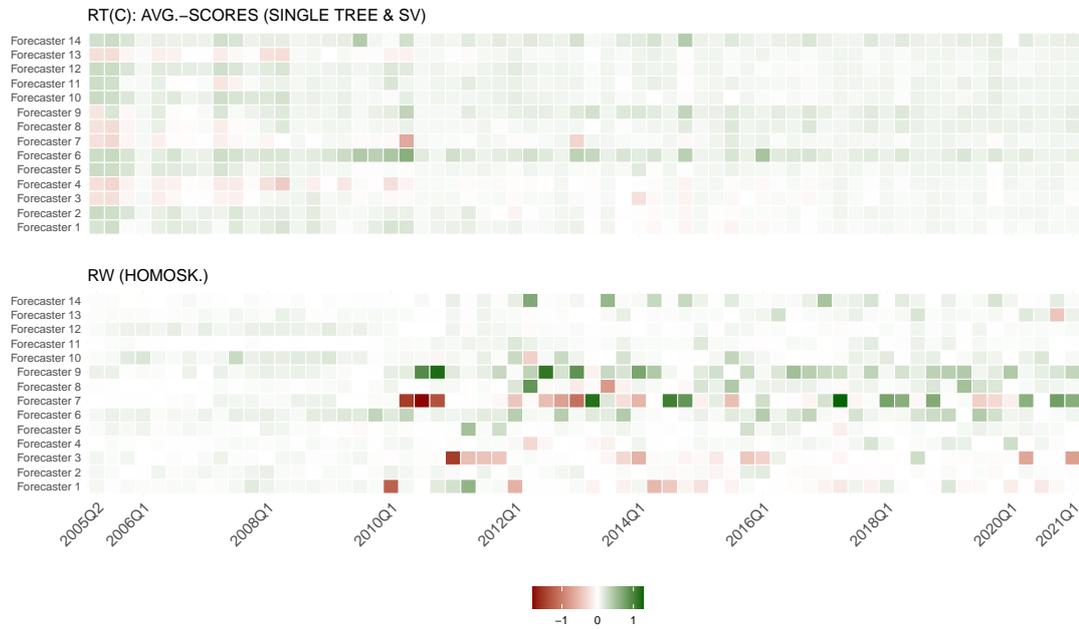
Figure B.2: Relative tail forecast accuracy: US inflation.



**Notes:** This figure shows two variants of quantile-weighted CRPS ratios, one focusing on the left tail and the other on the right tail. The gray-shaded entries give the actual scores of our benchmark (BPS-RW with homoskedastic error variances). Green-shaded entries refer to models that outperform the benchmark (with the forecast metric ratios below one), while red-shaded entries denote models that are outperformed by the benchmark (with the forecast metric ratios greater than one). The best-performing model specification by forecast metric is given in bold. Asterisks indicate statistical significance of the Diebold and Mariano (1995) test, which assumes equal forecast performance for each model relative to the benchmark, at the 1 (\*\*\*) , 5 (\*\*), and 10 (\*) percent significance levels.

## B.2.2 Combination Weights

Figure B.3: Combination weights over the evaluation sample: EA GDP growth



**Notes:** This figure shows the posterior median of the combination weights,  $(\gamma_j + \beta_{jt+h})$ , for each of the 14 SPF forecasters. Green (red)-shaded cells indicate that calibration parameters are above (below) zero. The top panel corresponds to our preferred BPS-RT specification, while the bottom panel corresponds to the benchmark.

Appendix B. Chapter 3 Appendix

Figure B.4: Combination weights over the evaluation sample: One-quarter-ahead US inflation



**Notes:** This figure shows the posterior median of the one-quarter-ahead combination weights,  $(\gamma_j + \beta_{jt+h})$ , for each of the 56 ADL model variants. Green (red)-shaded cells indicate that weights are above (below) zero. The top panel corresponds to our preferred BPS-RT specification, while the bottom panel corresponds to the benchmark.

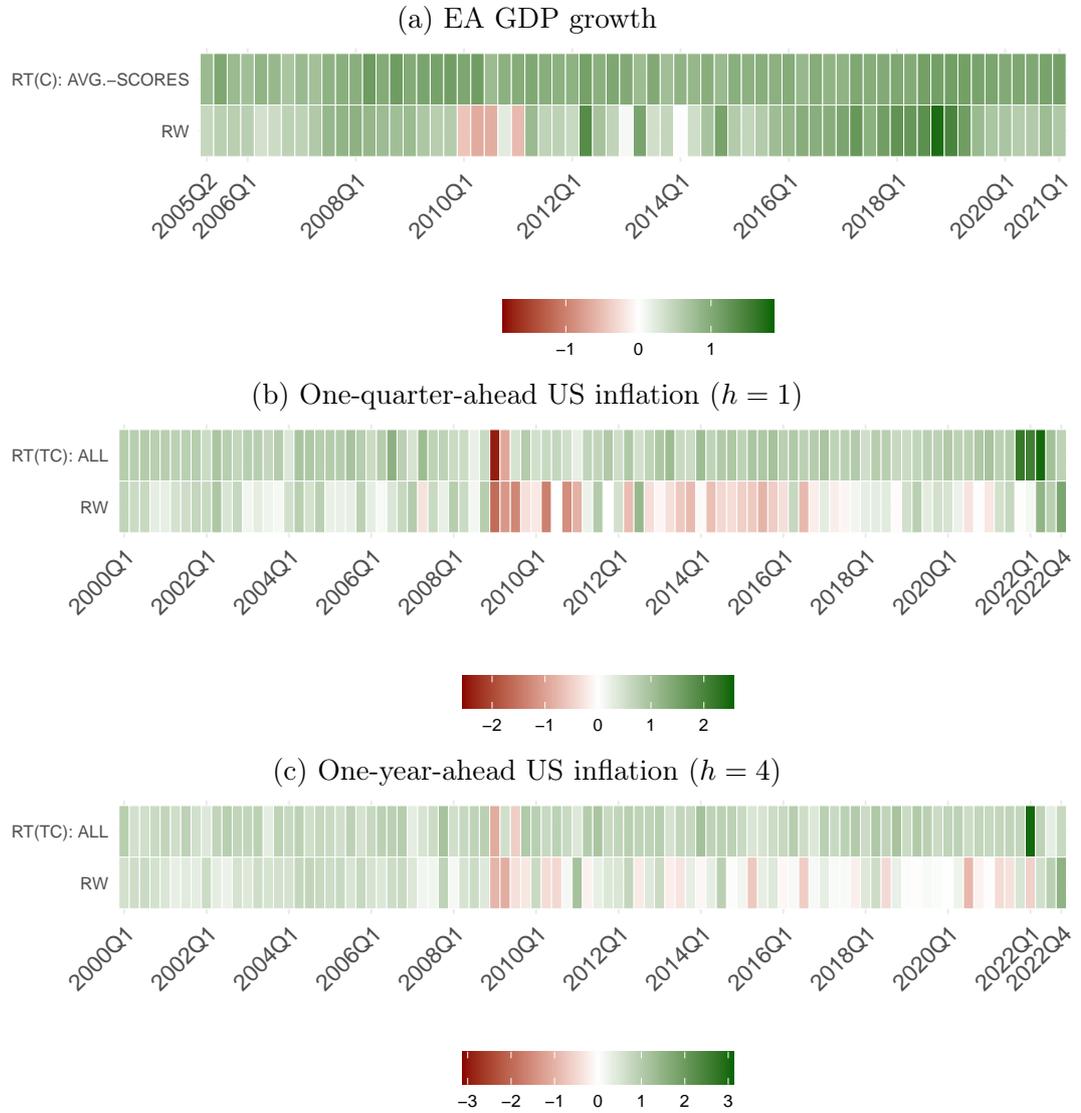
Appendix B. Chapter 3 Appendix

Figure B.5: Combination weights over the evaluation sample: One-year-ahead US inflation ( $h = 4$ )



**Notes:** This figure shows the posterior median of the one-year-ahead combination weights,  $(\gamma_j + \beta_{jt+h})$ , of the best-performing model parameters for each of the 56 ADL model variants. Green (red)-shaded cells indicate that weights are above (below) zero. The top panel corresponds to our preferred BPS-RT specification, while the bottom panel corresponds to the benchmark.

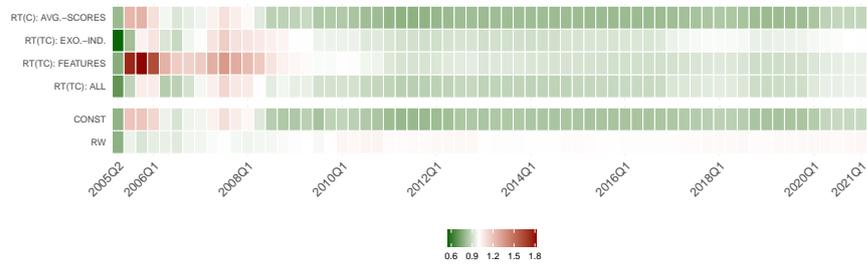
Figure B.6: Sum of combination weights over the evaluation sample



**Notes:** This figure shows the posterior median of the sum of the combination weights,  $\sum_{j=1}^J(\gamma_j + \beta_{jt+h})$ , for the models shown in Figures B.3, B.4 and B.5. Green (red)-shaded cells indicate that the overall sum of weights is above (below) zero for a specific evaluation period.

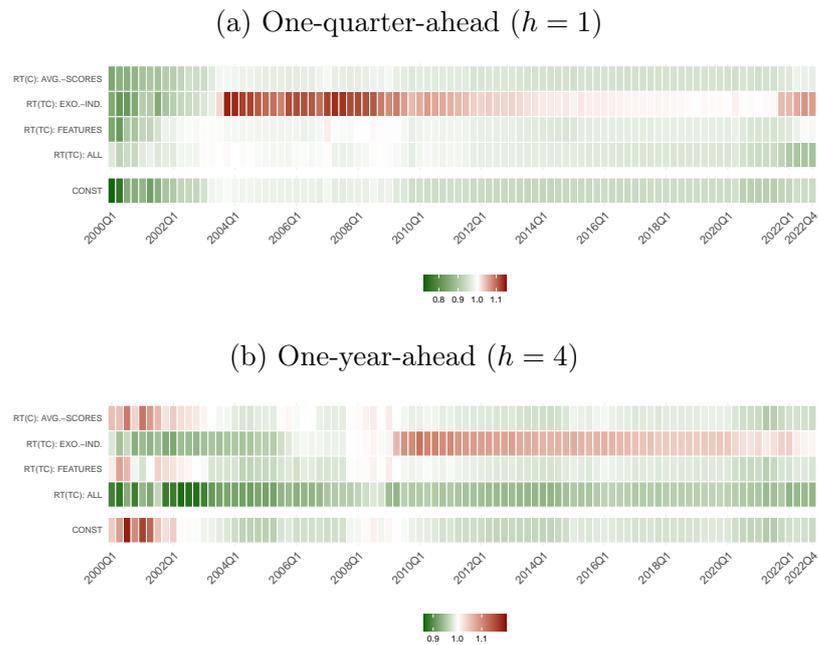
### B.2.3 Cumulative CRPS

Figure B.7: Forecast performance of single-tree specifications with stochastic volatility: EA GDP growth



**Notes:** This figure shows relative cumulative continuous ranked probability scores (CRPSs) over the full evaluation sample, which ranges from 2005Q2 to 2021Q1. The benchmark model is a TVP regression with a random walk evolution of parameters (BPS-RW) and homoskedastic error variances. Green-shaded entries indicate periods in which the respective model outperforms the benchmark (with the cumulative CRPS ratio below one), while red-shaded entries denote periods in which the respective model is outperformed by the benchmark (with the cumulative CRPS ratio greater than one). We refrain from showing the forecast performance over time for all models, but focus on the class of models that contains the best-performing specification in terms of CRPS, that is, all single-tree specifications with stochastic volatility.

Figure B.8: Forecast performance of single-tree specifications with homoskedastic error variances: US inflation

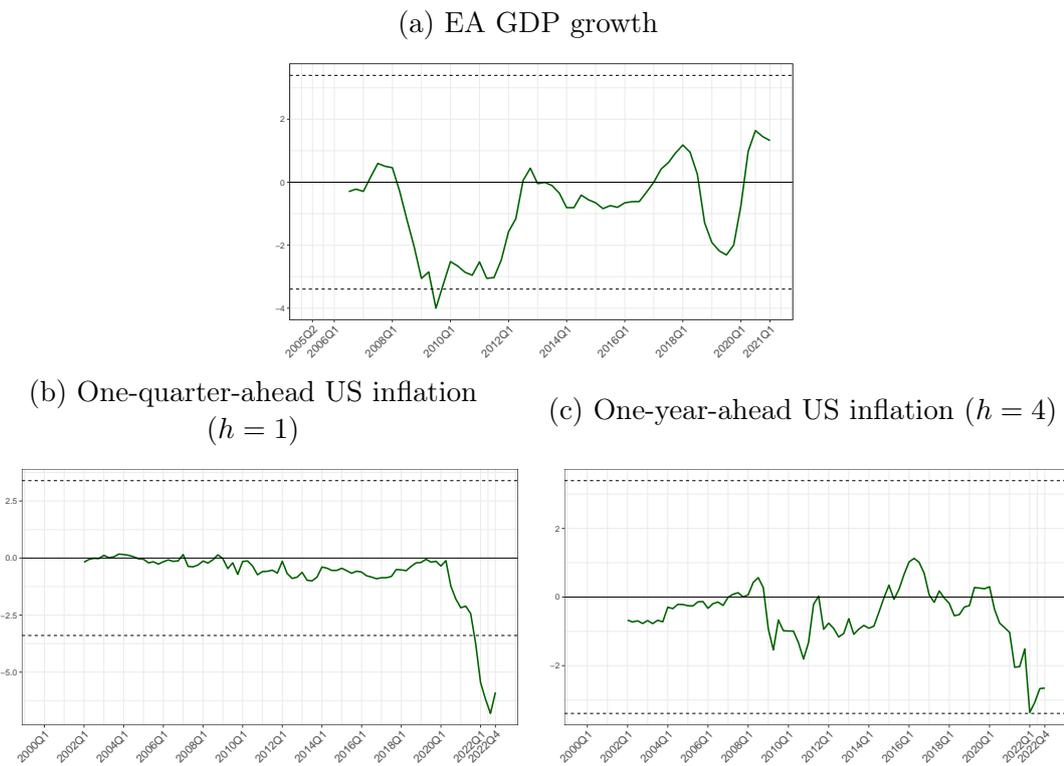


**Notes:** This figure shows average relative cumulative continuous ranked probability scores (CRPSs) over the evaluation sample, which ranges from 2000Q1 to 2022Q4. The benchmark model is a TVP regression with a random walk evolution of parameters (BPS-RW) and homoskedastic error variances. Green-shaded entries indicate periods in which the respective model outperforms the benchmark (with the cumulative CRPS ratio below one), while red-shaded entries denote periods in which the respective model is outperformed by the benchmark (with the cumulative CRPS ratio greater than one). We refrain from showing the forecast performance over time for all models, but focus on the class of models that forecast well in terms of CRPS, that is, all homoskedastic, single-tree specifications, the class containing the best-performing specification for the one-quarter-ahead ( $h = 1$ ) horizon.

### B.2.4 Giacomini and Rossi (2010) Fluctuation Test Statistic

We focus on evaluating the BPS-RT specifications with the best overall forecast performance. As seen from Figures 3.2 and 3.3, in the EA-GDP application, this is the single-tree specification with SV using average scores as effect modifiers. For the US inflation application, this is the homoskedastic single-tree specification with the full set of weight modifiers.

Figure B.9: Evolution of the Giacomini and Rossi (2010) fluctuation test statistic

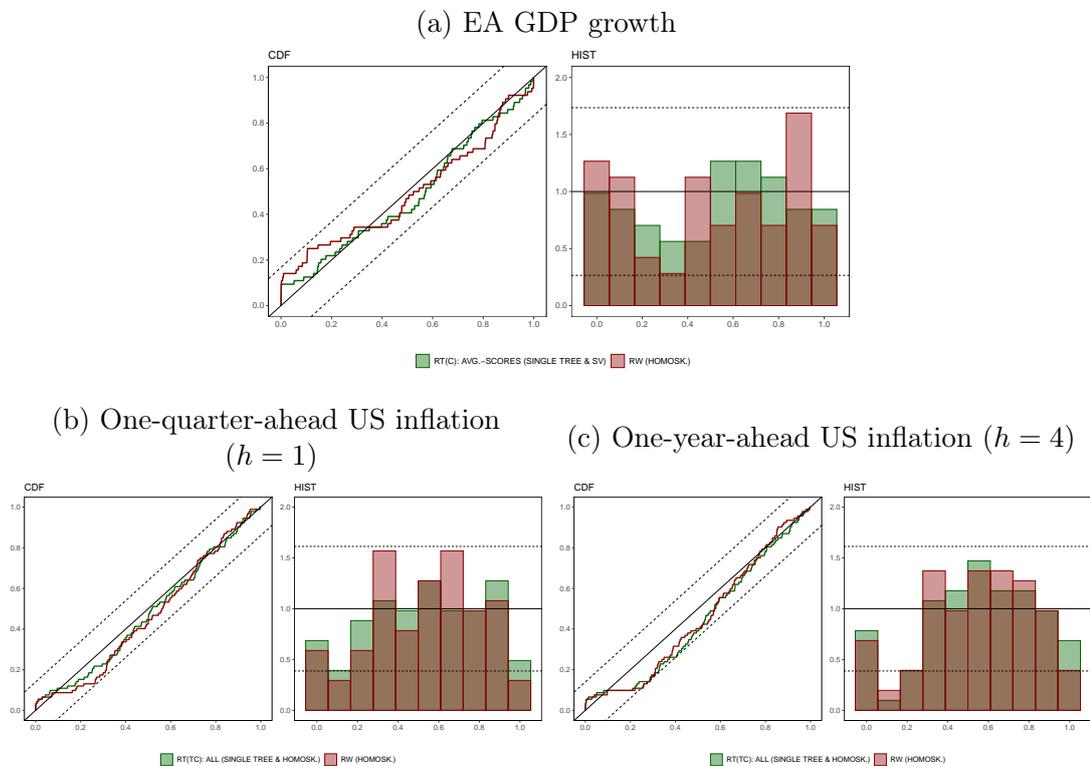


**Notes:** This figure shows the evolution of the Giacomini and Rossi (2010) fluctuation test statistic over time. The green solid line represents the test statistic, the black solid line marks the zero line, and the black dashed lines indicate the respective 95% confidence bands. To compute this period-specific test statistic, we use local relative continuous ranked probability scores (CRPSs) between the preferred BPS-RT specification and the benchmark (homoskedastic BPS-RW) over a rolling window comprising 10% of the evaluation sample. In panel (a), this implies that the rolling window is based on five observations (with the initial value of the test statistic available in 2006Q3), while in panels (b) and (c), this implies the rolling window is based on eight observations (with the initial value of the test statistic available in 2002Q1).

### B.2.5 Probability Integral Transforms (PITs)

We focus on evaluating the BPS-RT specifications with the best overall forecast performance. As seen from Figures 3.2 and 3.3, in the EA-GDP application, this is the single tree specification with SV using average scores as effect modifiers. For the US inflation application, this is the homoskedastic single tree specification with the full set of weight modifiers.

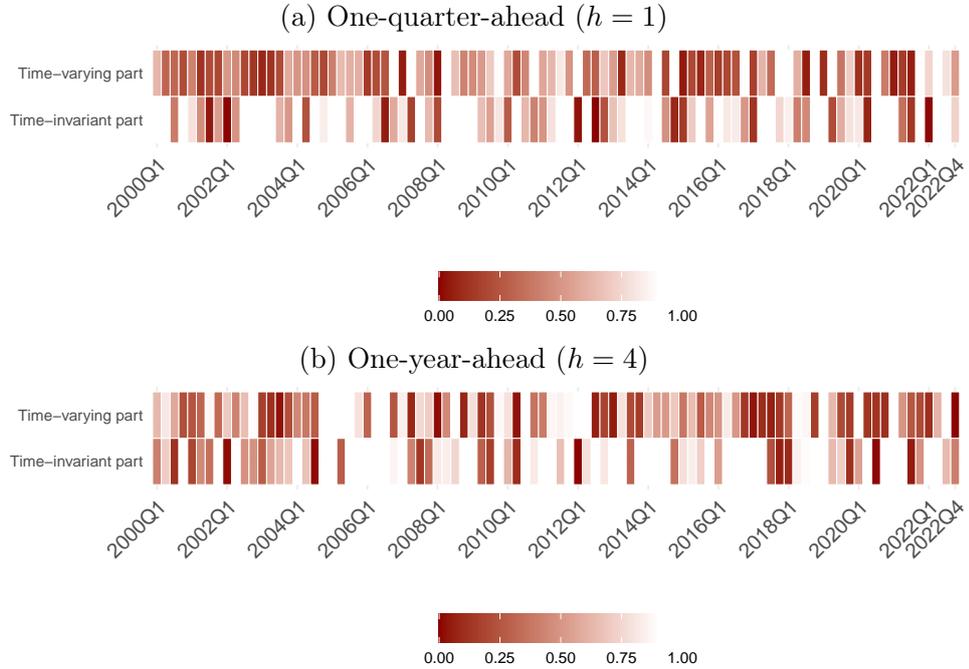
Figure B.10: Evaluating model calibration using probability integral transforms (PITs)



**Notes:** This figure shows the empirical cumulative density function of the PITs in the left panels and the histogram of the PITs in the right panels. A correctly specified model has PITs that are standard uniformly distributed. Such a specification is denoted by the black solid lines, with the black dashed lines denoting the respective 95% confidence bands (see Rossi and Sekhposyan, 2019). The preferred BPS-RT specification is shown in green, while the benchmark is indicated in red.

### B.2.6 Measuring the Degree of Shrinkage

Figure B.11: Overall degree of shrinkage toward the prior mean for US inflation.



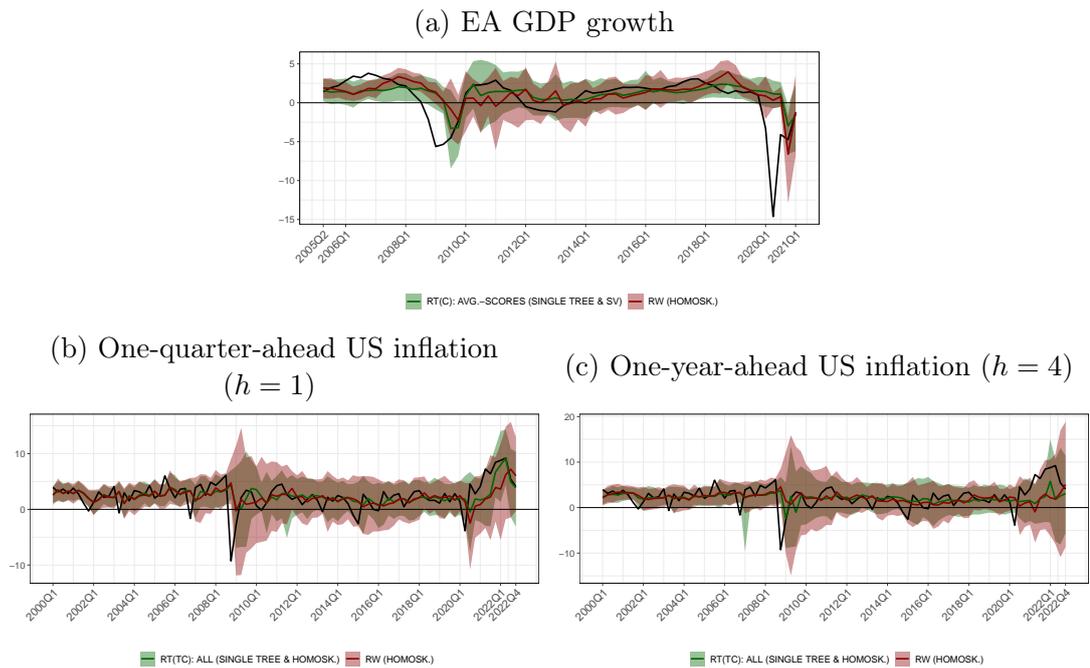
**Notes:** This figure shows the evolution of the degree of shrinkage measure over time. For each period in the evaluation sample, this measure is computed for our preferred specification (homoskedastic BPS-RT(TC): ALL with a single tree) as the ratio between the variation explained of the prior mean and the total variation of the respective coefficient part. This measure is bounded between zero and one. Values close to zero suggest that idiosyncratic deviations of coefficients (via the state innovation variances) dominate the prior mean (at least for some coefficients), while values close to one indicate that all coefficients are strongly pushed toward the prior mean in the respective period in the evaluation sample.

Figure B.11 shows a measure for overall shrinkage for both the time-invariant part ( $\gamma$ ) and the time-varying part ( $\beta_t$ ) of the weights. This measure effectively summarizes the overall variation explained by the prior mean as part of the overall variation in coefficients. It thus allows us to assess the relative importance of idiosyncratic innovations to the coefficients (i.e., innovations to the state equation) compared to the prior mean, and thus serves to quantify the overall degree of shrinkage by resembling something like a “joint”  $R^2$ -type of measure, which is bounded between zero and one. In each of the state equations, the target variables are either the constant coefficients  $\gamma$  or the

stacked time-varying coefficients  $\beta_t$  (for,  $t = 1, \dots, T$ ). In such a hierarchical model, the prior mean can then be treated as the conditional mean (i.e., the fit), while the state innovations (i.e., the shocks) are mainly driven by the state innovation (or prior) variances. In the following, a low joint  $R^2$  suggests that the state innovation variances play a significant role (at least for some of the coefficients), whereas a high joint  $R^2$  suggests that the coefficients are heavily shrunk toward the prior mean. It is worth noting that in recessions, the  $R^2$  is typically essentially zero and thus the prior means are less informative in these periods and random innovations to the state equations provide/add more model flexibility, which is required/necessary in these highly volatile periods.

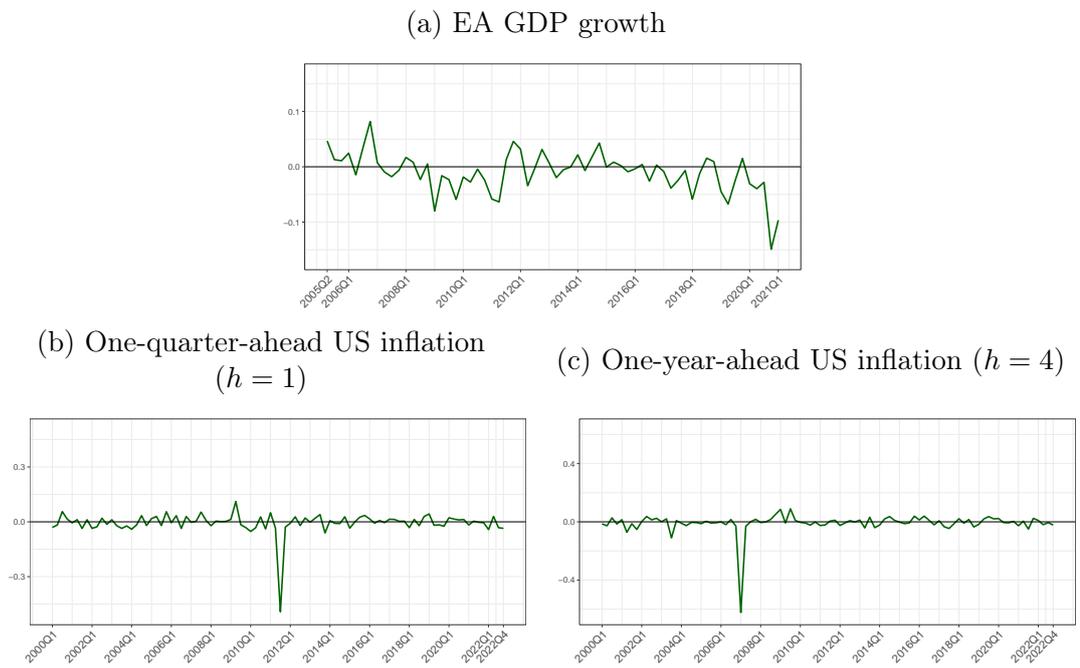
### B.2.7 Combined Predictive Densities and Their Skewness

Figure B.12: BPS-RT and BPS-RW predictive densities



**Notes:** This figure displays the corresponding predictive densities. The colored shaded areas and the colored solid lines represent the 90% confidence interval and the posterior median, respectively. The preferred BPS-RT specification is shown in green, while the benchmark is indicated in red. The black solid line in both panels refers to the respective realization.

Figure B.13: Evolution of a quantile-based skewness measure for the BPS-RT predictive densities



**Notes:** This figure shows the evolution of a quantile-based skewness measure for predictive densities in our preferred BPS-RT specification, as shown in Figure B.12. The quantile-based skewness measure is defined as  $((q_{95\%} - q_{50\%}) - (q_{50\%} - q_{5\%})) / (q_{95\%} - q_{5\%})$ , where  $q_{5\%}$ ,  $q_{50\%}$ , and  $q_{95\%}$  represent the 5<sup>th</sup>, 50<sup>th</sup>, and 95<sup>th</sup> percentiles of the predictive densities, respectively. The green solid line represents the computed skewness measure, while the black solid line marks the zero line.

## Appendix C

### Chapter 4 Appendix

## C.1 Evaluation of Tilting Vectors

Evaluation of tilting vectors  $\boldsymbol{\tau}(\mathbf{x})$  involves Monte Carlo integration and numerical optimization, briefly summarized here. This follows and customizes the general results in Tallman and West (2023, section 4.4).

The vector  $\boldsymbol{\tau}(\mathbf{x})$  is implicitly defined to satisfy  $E_f[\mathbf{s}_j(\mathbf{y}, \mathbf{x}_j)|\mathbf{x}] = \mathbf{m}_f$  where  $\mathbf{m}_f$  is the chosen target score and expectation is with respect to

$$f(\mathbf{y}, \mathcal{M}_j) = \tilde{\pi}_j(\mathbf{x})f_j(\mathbf{y}|\mathbf{x}, \mathcal{M}_j) = k(\mathbf{x})\pi_j(\mathbf{x})\alpha_j(\mathbf{y}|\mathbf{x})p_j(\mathbf{y}|\mathbf{x}, \mathcal{M}_j), \quad j = 0:J,$$

with components given in eqn. (4.3) in the main text. With  $\alpha_j(\mathbf{y}|\mathbf{x})$  of the exponential form in eqn. (4.4) in the main text, numerical optimization aims to solve

$$k(\mathbf{x}) \sum_{j=0:J} \int_{\mathbf{y}} \pi_j(\mathbf{x}) e^{\boldsymbol{\tau}(\mathbf{x})' \mathbf{s}_j(\mathbf{y}, \mathbf{x}_j)} p_j(\mathbf{y}|\mathbf{x}, \mathcal{M}_j) d\mathbf{y} - \mathbf{m}_f = \mathbf{0} \quad (\text{B.1})$$

for  $\boldsymbol{\tau}(\mathbf{x})$ . The normalizing term  $k(\mathbf{x})$  is defined via  $k(\mathbf{x})^{-1} = \sum_{j=0:J} \pi_j(\mathbf{x}) a_j(\mathbf{x})$  where

$$a_j(\mathbf{x}) = \int_{\mathbf{y}} e^{\boldsymbol{\tau}(\mathbf{x})' \mathbf{s}_j(\mathbf{y}, \mathbf{x}_j)} p_j(\mathbf{y}|\mathbf{x}, \mathcal{M}_j) d\mathbf{y}, \quad j = 0:J. \quad (\text{B.2})$$

At any given  $\mathbf{x}$ , the integrals in equations (B.1,B.2) are approximated using Monte Carlo integration based on random samples from each of the  $p_j(\mathbf{y}|\mathbf{x}, \mathcal{M}_j)$ . This gives the ingredients for numerical search over  $\boldsymbol{\tau}(\mathbf{x})$  using derivative-free optimization methods such as those used in the current paper.

For completeness, we note that the same approach applies where gradient-based optimization is considered. Newton-Raphson and allied methods will require the vector derivative  $\dot{a}_j(\mathbf{x})$  of  $a_j(\mathbf{x})$  with respect to  $\boldsymbol{\tau}(\mathbf{x})$ , and may also require the matrix of second

derivatives  $\ddot{a}_j(\mathbf{x})$ . These are given by

$$\begin{aligned}\dot{a}_j(\mathbf{x}) &= \int_{\mathbf{y}} \mathbf{s}_j(\mathbf{y}, \mathbf{x}_j) e^{\boldsymbol{\tau}(\mathbf{x})' \mathbf{s}_j(\mathbf{y}, \mathbf{x}_j)} p_j(\mathbf{y}|\mathbf{x}, \mathcal{M}_j) d\mathbf{y}, \\ \ddot{a}_j(\mathbf{x}) &= \int_{\mathbf{y}} \mathbf{s}_j(\mathbf{y}, \mathbf{x}_j) \mathbf{s}_j(\mathbf{y}, \mathbf{x}_j)' e^{\boldsymbol{\tau}(\mathbf{x})' \mathbf{s}_j(\mathbf{y}, \mathbf{x}_j)} p_j(\mathbf{y}|\mathbf{x}, \mathcal{M}_j) d\mathbf{y}.\end{aligned}$$

Hence the Monte Carlo integration approach used to evaluate the  $a_j(\mathbf{y})$  also delivers approximations to these derivatives.

## C.2 Conditional Forecasts

This section discusses the construction of conditional forecasting in VAR models, following Chan *et al.* (2023) and discussion in Antolin-Diaz *et al.* (2021).

### C.2.1 Observational Constraints

The general theory for constraining on exact or uncertain linear constraints in multivariate normal forecast distributions (e.g. West and Harrison, 1997, sect. 16.3.2) underlies these results used in that paper and here in the specific case of VAR models.

The  $n \times 1$  vector time series  $\mathbf{y}_t = (y_{t1}, \dots, y_{tn})'$  follows a VAR( $p$ ) in structural form

$$\mathbf{A}_0 \mathbf{y}_t = \mathbf{a} + \mathbf{A}_1 y_{t-1} + \dots + \mathbf{A}_p y_{t-p} + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim N(\mathbf{0}_n, \mathbf{I}_n) \quad (\text{B.3})$$

with the following terms:  $\mathbf{a}$  is a  $n \times 1$  vector of intercepts,  $\mathbf{A}_1, \dots, \mathbf{A}_p$  are the  $n \times n$  VAR coefficient matrices,  $\mathbf{A}_0$  is a full-rank contemporaneous impact matrix,  $\mathbf{0}_n$  is a  $n \times 1$  zero vector and  $\mathbf{I}_n$  is the  $n \times n$  identity matrix.

Given the history of observations  $\{\mathbf{y}_{1-p}, \dots, \mathbf{y}_T\}$  at any time  $T$ , unconditional forecasting of the  $n$  variables for the next  $h$  periods, namely the  $nh \times 1$  vector  $\mathbf{y}_{T+1:T+h} = (\mathbf{y}'_{T+1}, \dots, \mathbf{y}'_{T+h})'$ , is based on the implied form

$$\mathbf{H} \mathbf{y}_{T+1:T+h} = \mathbf{c} + \boldsymbol{\varepsilon}_{T+1:T+h}, \quad \boldsymbol{\varepsilon}_{T+1:T+h} \sim N(\mathbf{0}_{nh}, \mathbf{I}_{nh}) \quad (\text{B.4})$$

where

$$\mathbf{c} = \begin{bmatrix} a + \sum_{j=1}^p \mathbf{A}_j \mathbf{y}_{T+1-j} \\ a + \sum_{j=2}^p \mathbf{A}_j \mathbf{y}_{T+1-j} \\ \vdots \\ a + \mathbf{A}_p \mathbf{y}_T \\ a \\ \vdots \\ a \end{bmatrix}, \quad \mathbf{H} = \begin{bmatrix} \mathbf{A}_0 & \mathbf{0}_{n \times n} & \cdots & \cdots & \cdots & \cdots & \cdots & \mathbf{0}_{n \times n} \\ -\mathbf{A}_1 & \mathbf{A}_0 & \mathbf{0}_{n \times n} & \cdots & \cdots & \cdots & \cdots & \mathbf{0}_{n \times n} \\ -\mathbf{A}_2 & -\mathbf{A}_1 & \mathbf{A}_0 & \mathbf{0}_{n \times n} & \cdots & \cdots & \cdots & \mathbf{0}_{n \times n} \\ \vdots & \ddots & \ddots & \ddots & \cdots & \cdots & \cdots & \vdots \\ -\mathbf{A}_p & \cdots & \cdots & -\mathbf{A}_1 & \mathbf{A}_0 & \mathbf{0}_{n \times n} & \cdots & \mathbf{0}_{n \times n} \\ \mathbf{0}_{n \times n} & \cdots & \cdots & \ddots & \ddots & \ddots & \cdots & \vdots \\ \vdots & \cdots & \cdots & \cdots & \ddots & \ddots & \ddots & \vdots \\ \mathbf{0}_{n \times n} & \cdots & \mathbf{0}_{n \times n} & -\mathbf{A}_p & \cdots & \cdots & -\mathbf{A}_1 & \mathbf{A}_0 \end{bmatrix}$$

with  $\mathbf{0}_{n \times n}$  the  $n \times n$  zero matrix. Since  $\mathbf{A}_0$  is full rank and  $|\mathbf{H}| = |\mathbf{A}_0|^h \neq 0$ , then  $\mathbf{H}$  is non-singular and

$$\mathbf{y}_{T+1:T+h} \sim N(\mathbf{m}, \mathbf{M}) \quad \text{with moments} \quad \mathbf{m} = \mathbf{H}^{-1} \mathbf{c} \quad \text{and} \quad \mathbf{M} = (\mathbf{H}' \mathbf{H})^{-1}. \quad (\text{B.5})$$

Now consider a set of  $r < nh$  linear restrictions on the path of future observables, namely

$$\mathbf{R} \mathbf{y}_{T+1:T+h} \sim N(\mathbf{r}, \mathbf{\Omega}) \quad (\text{B.6})$$

where  $\mathbf{R}$  is a chosen  $r \times nh$  constant matrix with full row rank,  $\mathbf{r}$  is an  $r \times 1$  vector mean and  $\mathbf{\Omega}$  is the corresponding  $r \times r$  variance matrix related to the restrictions. This can be regarded as a set of uncertain constraints  $\mathbf{r}$  when  $\mathbf{\Omega}$  is non-zero, or exact constraints  $\mathbf{R} \mathbf{y}_{T+1:T+h} = \mathbf{r}$  in the case  $\mathbf{\Omega} = \mathbf{0}$ .

Conditioning the forecast distribution of eqn. (B.5) on this constraint information yields the updated distribution below; details follow from West and Harrison (sect. 16.3.2 1997) and Chan *et al.* (2023). The updated distribution for the path is

$$\mathbf{y}_{T+1:T+h} | \mathbf{R}, \mathbf{r}, \mathbf{\Omega} \sim N(\mathbf{m}^*, \mathbf{M}^*),$$

with

$$\mathbf{m}^* = \mathbf{m} + \mathbf{A}(\mathbf{r} - \mathbf{R}\mathbf{m}) \quad \text{and} \quad \mathbf{M}^* = \mathbf{M} + \mathbf{A}(\mathbf{\Omega} - \mathbf{R}\mathbf{M}\mathbf{R}')\mathbf{A}' \quad \text{where} \quad \mathbf{A} = \mathbf{M}\mathbf{R}'(\mathbf{R}\mathbf{M}\mathbf{R}')^{-1}.$$

The special case of exact constraints  $\mathbf{R}\mathbf{y}_{T+1:T+h} = \mathbf{r}$  has constraint uncertainty matrix  $\mathbf{\Omega} = \mathbf{0}$  so that  $\mathbf{M}^*$  reduces to  $\mathbf{M}^* = \mathbf{M} - \mathbf{A}\mathbf{R}\mathbf{M}\mathbf{R}'\mathbf{A}'$ . This is relevant in applications where it is justifiable to assume exact constraints. In our policy setting, the targeted constraint vector  $\mathbf{r}$  is just that, a target policy path, for example, so that there will always be some level of uncertainty. The starting point is uncertainty represented by the variance matrix  $V(\mathbf{R}\mathbf{y}_{T+1:T+h}) = \mathbf{R}\mathbf{M}\mathbf{R}'$ . As argued in Antolin-Diaz *et al.* (2021), adopting  $\mathbf{\Omega} = \mathbf{R}\mathbf{M}\mathbf{R}'$  represents a position that admits relevant and *conservative* levels of such uncertainty. This is used in the application of our paper as we constrain on candidate values of the policy path in considering conditional forecasts, hedged with uncertainty.

## C.2.2 Constraints on Structural Shocks

The above development implicitly involves constraints imposed on all of the structural shocks of the model. In conditional forecasting exercises, it is often necessary to focus on only specific structural shocks set at restricted values (exact or uncertain), to obtain finer control. For example, a forecast conditional on an increasing policy rate path might result in an increase in inflation (i.e., the “price puzzle”) as the forecast is driven by reduced form shocks which are a mix of the structural shocks (i.e., demand, supply, monetary). For further discussion, see Antolin-Diaz *et al.* (2021). Due to these issues, we impose structural restrictions so that changes in the policy rate are driven by monetary policy shocks (i.e., decisions from the Central Bank) rather than as reactions to other shocks. This can be done by considering the restrictions

$$\mathbf{W}\boldsymbol{\varepsilon}_{T+1:T+h} \sim N(\mathbf{w}, \boldsymbol{\Psi}) \tag{B.7}$$

where  $\mathbf{W}$  is a  $w \times nh$  full rank-selection matrix,  $\mathbf{w}$  is a  $w \times 1$  vector of constants and  $\Psi$  is a  $w \times w$  covariance matrix. Substituting eqn. (B.4) into eqn. (B.7) for  $\varepsilon_{T+1:T+h}$  results in

$$\mathbf{W}\mathbf{H}\mathbf{y}_{T+1:T+h} \sim N(\mathbf{W}\mathbf{c}, \mathbf{I}_w) \quad (\text{B.8})$$

Then combining eqn. (B.8) with eqn. (B.6) yields

$$\underbrace{\begin{bmatrix} \mathbf{R}_0 \\ \mathbf{W}\mathbf{H} \end{bmatrix}}_{\mathbf{R}} \mathbf{y}_{T+1:T+h} \sim N \left( \underbrace{\begin{bmatrix} \mathbf{r}_0 \\ \mathbf{W}\mathbf{c} \end{bmatrix}}_{\mathbf{r}}, \underbrace{\begin{bmatrix} \mathbf{\Omega}_0 & \mathbf{0}_{r_0 \times w} \\ \mathbf{0}_{w \times r_0} & \mathbf{I}_w \end{bmatrix}}_{\mathbf{\Omega}} \right) \quad (\text{B.9})$$

where  $\mathbf{R}_0$  is a selection matrix for observable restrictions,  $\mathbf{r}_0$  is a vector of restrictions on observables, and  $\mathbf{\Omega}_0$  is a matrix of covariance restrictions on observables. Thus, by setting  $\mathbf{R}$ ,  $\mathbf{r}$  and  $\mathbf{\Omega}$  as indicated in eqn. (B.8), the structural restrictions can be regarded as a specific case of eqn. (B.6). In our applications, we set  $\mathbf{r}_0 = \mathbf{x}$  and  $\mathbf{W} = \mathbf{0}$  for all  $\varepsilon_{T+1:T+h}$  except those associated with the monetary policy shock—those are kept unrestricted. This amounts to conditioning on a proposed decision vector  $\mathbf{x}$  and assuming it is driven only by monetary policy shocks. Other conditioning assumptions are possible. For example, in addition to assuming the driving shock is monetary policy one could assume a vector of positive/negative demand shocks representing different future scenarios.

# Bibliography

- AASTVEIT KA, CROSS JL, AND VAN DIJK HK (2023), “Quantifying time-varying forecast uncertainty and risk for the real price of oil,” *Journal of Business & Economic Statistics* **41**(2), 523–537.
- AASTVEIT KA, GERDRUP K, AND JORE AS (2011), “Short-term forecasting of GDP and inflation in real-time: Norges Bank’s system for averaging models,” *Norges Bank Staff Memo* (9).
- AASTVEIT KA, GERDRUP KR, JORE AS, AND THORSRUD LA (2014), “Nowcasting GDP in real time: A density combination approach,” *Journal of Business & Economic Statistics* **32**(1), 48–68.
- AASTVEIT KA, MITCHELL J, RAVAZZOLO F, AND VAN DIJK HK (2019), “The evolution of forecast density combinations in economics,” *Oxford Research Encyclopedia of Economics and Finance* DOI: 10.1093/acrefore/9780190625979.013.381.
- AASTVEIT KA, RAVAZZOLO F, AND VAN DIJK HK (2018), “Combined density nowcasting in an uncertain economic environment,” *Journal of Business & Economic Statistics* **36**(1), 131–145.
- ADRIAN T, BOYARCHENKO N, AND GIANNONE D (2019), “Vulnerable growth,” *American Economic Review* **109**(4), 1263–1289.
- ANTOLIN-DIAZ J, PETRELLA I, AND RUBIO-RAMIREZ JF (2021), “Structural scenario analysis with SVARs,” *Journal of Monetary Economics* **117**, 798–815.

## Bibliography

- BACHE IW, MITCHELL J, RAVAZZOLO F, AND VAHEY SP (2009), “Macro modelling with many models,” Technical Report 2009/15, Norges Bank.
- BAKER S, BLOOM N, AND DAVIS S (2016), “Measuring economic policy uncertainty,” *The Quarterly Journal of Economics* **131**(4), 1593–1636.
- BASSETTI F, CASARIN R, AND DEL NEGRO M (2023), “Inference on probabilistic surveys in macroeconomics with an application to the evolution of uncertainty in the Survey of Professional Forecasters during the COVID pandemic,” in R BACHMANN, G TOPA, AND W VAN DER KLAAUW (eds.) “Handbook of Economic Expectations,” 443–476, Elsevier.
- BASSETTI F, CASARIN R, AND RAVAZZOLO F (2018), “Bayesian nonparametric calibration and combination of predictive distributions,” *Journal of the American Statistical Association* **113**(522), 675–685.
- BATES JM, AND GRANGER CWJ (1969), “The Combination of Forecasts,” *Journal of the Operational Research Society* **20**(4), 451–468.
- BELMONTE MAG, KOOP G, AND KOROBILIS D (2014), “Hierarchical shrinkage in time-varying parameter models,” *Journal of Forecasting* **33**(1), 80–94, publisher: John Wiley & Sons, Ltd.
- BERNACIAK D, AND GRIFFIN JE (2024), “A loss discounting framework for model averaging and selection in time series models,” *International Journal of Forecasting* [Doi.org/10.1016/j.ijforecast.2024.03.001](https://doi.org/10.1016/j.ijforecast.2024.03.001).
- BILLIO M, CASARIN R, RAVAZZOLO F, AND VAN DIJK HK (2013), “Time-varying combinations of predictive densities using nonlinear filtering,” *Journal of Econometrics* **177**(2), 213–232.
- BITTO A, AND FRÜHWIRTH-SCHNATTER S (2019), “Achieving shrinkage in a time-varying parameter model framework,” *Journal of Econometrics* **210**(1), 75–97.

## Bibliography

- BJØRNLAND HC, GERDRUP K, JORE AS, SMITH C, AND THORSRUD LA (2012), “Does forecast combination improve Norges Bank inflation forecasts?” *Oxford Bulletin of Economics and Statistics* **74**(2), 163–179.
- BROWN PJ, AND GRIFFIN JE (2010), “Inference with normal-gamma prior distributions in regression problems,” *Bayesian Analysis* **5**(1), 171–188, publisher: International Society for Bayesian Analysis.
- CADONNA A, FRÜHWIRTH-SCHNATTER S, AND KNAUS P (2020), “Triple the Gamma—A Unifying Shrinkage Prior for Variance and Variable Selection in Sparse State Space and TVP Models,” *Econometrics* **8**(2).
- ČAPEK J, CUARESMA JC, HAUZENBERGER N, AND REICHEL V (2023), “Macroeconomic forecasting in the euro area using predictive combinations of DSGE models,” *International Journal of Forecasting* **39**(4), 1820–1838.
- CARTER C, AND KOHN R (1994), “On Gibbs sampling for state space models,” *Biometrika* **81**(3), 541–553.
- CARVALHO CM, POLSON NG, AND SCOTT JG (2010), “The horseshoe estimator for sparse signals,” *Biometrika* **97**(2), 465–480.
- CASARIN R, GRASSI S, RAVAZZOLLO F, AND VAN DIJK HK (2019), “Forecast density combinations with dynamic learning for large data sets in economics and finance,” SSRN Scholarly Paper 3363556, Social Science Research Network, Rochester, NY.
- CHAN JC, AND JELIAZKOV I (2009), “Efficient simulation and integrated likelihood estimation in state space models,” *International Journal of Mathematical Modelling and Numerical Optimisation* **1**(1-2), 101–120.
- CHAN JCC (2022), “Asymmetric conjugate priors for large Bayesian VARs,” *Quantitative Economics* **13**, 1145–1169.

## Bibliography

- CHAN JCC, PETTENUZZO D, POON A, AND ZHU D (2023), “Conditional Forecasts in Large Bayesian VARs with Multiple Soft and Hard Constraints,” *SSRN Electronic Journal* [Doi.org/10.2139/ssrn.4358152](https://doi.org/10.2139/ssrn.4358152).
- CHERNIS T (2023), “Combining Large Numbers of Density Predictions with Bayesian Predictive Synthesis,” *Studies in Nonlinear Dynamics & Econometrics* .
- (2024), “Combining Large Numbers of Density Predictions with Bayesian Predictive Synthesis,” *Studies in Nonlinear Dynamics Econometrics* **28**(2), 293–317.
- CHERNIS T, HAUZENBERGER N, HUBER F, KOOP G, AND MITCHELL J (2023), “Predictive Density Combination Using a Tree-Based Synthesis Function,” *arXiv preprint arXiv:2311.12671* .
- CHERNIS T, KOOP G, TALLMAN E, AND WEST M (2024), “Decision synthesis in monetary policy,” *arXiv preprint arXiv:2406.03321* .
- CHERNIS T, AND SEKKEL R (2017), “A dynamic factor model for nowcasting Canadian GDP growth,” *Empirical Economics* **53**(1), 217–234.
- (2018), “Nowcasting Canadian economic activity in an uncertain environment,” Discussion Paper, Bank of Canada.
- CHERNIS T, AND WEBLEY T (2022), “Nowcasting Canadian GDP with density combinations,” Technical report, Bank of Canada, [doi.org/10.34989/sdp-2022-12](https://doi.org/10.34989/sdp-2022-12).
- CHIPMAN HA, GEORGE EI, AND MCCULLOCH RE (1998), “Bayesian CART model search,” *Journal of the American Statistical Association* **93**(443), 935–948.
- (2010), “BART: Bayesian additive regression trees,” *The Annals of Applied Statistics* **4**(1), 266–298.
- CLARK TE (2011), “Real-time density forecasts from Bayesian Vector Autoregressions With stochastic volatility,” *Journal of Business & Economic Statistics* **29**(3), 327–341.

## Bibliography

- CLARK TE, HUBER F, KOOP G, MARCELLINO M, AND PFARRHOFER M (2023), “Tail forecasting with multivariate Bayesian additive regression trees,” *International Economic Review* **64**(3), 979–1022.
- COLETTI D, AND MURCHISON S (2002), “Models in policy-making,” *Bank of Canada Review* **2002**(Spring), 19–26.
- CONFLITTI C, DE MOL C, AND GIANNONE D (2015), “Optimal combination of survey forecasts,” *International Journal of Forecasting* **31**, 1096–1103.
- CROSS JL, HOU C, AND POON A (2020), “Macroeconomic forecasting with large Bayesian VARs: Global-local priors and the illusion of sparsity,” *International Journal of Forecasting* **36**(3), 899–915.
- DEL NEGRO M, HASEGAWA RB, AND SCHORFHEIDE F (2016), “Dynamic prediction pools: An investigation of financial frictions and forecasting performance,” *Journal of Econometrics* **192**(2), 391–405.
- DESHPANDE SK, BAI R, BALOCCHI C, STARLING JE, AND WEISS J (2020), “VCBART: Bayesian trees for varying coefficients,” Technical report, arXiv preprint arXiv:2003.06416.
- DIEBOLD FX, GUNTHER TA, AND TAY AS (1998), “Evaluating density forecasts with applications to financial risk Management,” *International Economic Review* **39**(4), 863–883.
- DIEBOLD FX, AND MARIANO RS (1995), “Comparing Predictive Accuracy,” *Journal of Business & Economic Statistics* **13**(3), 253–263.
- DIEBOLD FX, SHIN M, AND ZHANG B (2023), “On the aggregation of probability assessments: Regularized mixtures of predictive densities for Eurozone inflation and real interest rates,” *Journal of Econometrics* **237**, 105321.

## Bibliography

- ECB (2019), “Results of the third special questionnaire for participants in the ECB Survey of Professional Forecasters,” .
- FAVA B, AND LOPES HF (2021), “The illusion of the illusion of sparsity: An exercise in prior sensitivity,” *Brazilian Journal of Probability and Statistics* **35**(4), 699–720.
- FRÜHWIRTH-SCHNATTER S, AND WAGNER H (2010), “Stochastic model specification search for Gaussian and partial non-Gaussian state space models,” *Journal of Econometrics* **154**(1), 85–100.
- FURLANETTO F, RAVAZZOLO F, AND SARFERAZ S (2019), “Identification of financial factors in economic fluctuations,” *Economic Journal* **129**, 311–337.
- GARCÍA JA (2003), “An introduction to the ECB’s survey of professional forecasters,” Occasional Paper Series 8, European Central Bank.
- GEWEKE J (2010), *Complete and Incomplete Econometric Models*, Princeton: Princeton University Press.
- GEWEKE J, AND AMISANO G (2011), “Optimal prediction pools,” *Journal of Econometrics* **164**(1), 130–141.
- GIACOMINI R, AND ROSSI B (2010), “Forecast comparisons in unstable environments,” *Journal of Applied Econometrics* **25**(4), 595–620.
- GIANNONE D, LENZA M, AND PRIMICERI GE (2021), “Economic predictions with big data: The illusion of sparsity,” *Econometrica* **89**(5), 2409–2437, publisher: Econometric Society.
- GNEITING T, AND RAFTERY AE (2007), “Strictly proper scoring rules, prediction, and estimation,” *Journal of the American Statistical Association* **102**(477), 359–378.
- GNEITING T, AND RANJAN R (2011), “Comparing density forecasts using threshold- and quantile-Weighted scoring rules,” *Journal of Business & Economic Statistics* **29**(3), 411–422.

## Bibliography

- GOULET COULOMBE P (2024), “The macroeconomy as a random forest,” *Journal of Applied Econometrics* .
- GRUBER L, AND WEST M (2016), “GPU-accelerated Bayesian learning and forecasting in simultaneous graphical dynamic linear models,” *Bayesian Analysis* **11**, 125–149.
- GRUBER LF, AND WEST M (2017), “Bayesian forecasting and scalable multivariate volatility analysis using simultaneous graphical dynamic linear models,” *Econometrics and Statistics* **3**, 3–22.
- HALL SG, AND MITCHELL J (2007), “Combining density forecasts,” *International Journal of Forecasting* **23**(1), 1–13.
- HARTKOPF J (2022), “Gigrnd,” *MATLAB Central File Exchange*  
<https://ww2.mathworks.cn/matlabcentral/fileexchange/78805-gigrnd>.
- HAUZENBERGER N, HUBER F, KOOP G, AND MITCHELL J (2023), “Bayesian modeling of time-varying parameters using regression trees,” Federal Reserve Bank of Cleveland WP No. 23-05.
- HAUZENBERGER N, HUBER F, KOOP G, AND ONORANTE L (2022), “Fast and flexible Bayesian inference in time-varying parameter regression models,” *Journal of Business & Economic Statistics* **40**(4), 1904–1918.
- HUBER F, KOOP G, ONORANTE L, PFARRHOFER M, AND SCHREINER J (2023), “Nowcasting in a pandemic using non-parametric mixed frequency VARs,” *Journal of Econometrics* **232**(1), 52–69.
- HUBER F, AND ROSSINI L (2022), “Inference in Bayesian additive vector autoregressive tree models,” *The Annals of Applied Statistics* **16**(1), 104–123.
- HÖRMANN W, AND LEYDOLD J (2014), “Generating generalized inverse Gaussian random variates,” *Statistics and Computing* **24**(4), 547–557.

## Bibliography

- JIN X, MAHEU JM, AND YANG Q (2022), “Infinite Markov pooling of predictive distributions,” *Journal of Econometrics* **228**, 302–321.
- JOHNSON MC, AND WEST M (2022), “Bayesian predictive synthesis with outcome-dependent pools,” *Technical Report, Department of Statistical Science, Duke University* ArXiv:1803.01984.
- JORE AS, MITCHELL J, AND VAHEY SP (2010), “Combining forecast densities from VARs with uncertain instabilities,” *Journal of Applied Econometrics* **25**(4), 621–634.
- KAPETANIOS G, LABHARD V, AND PRICE S (2008), “Forecast combination and the Bank of England’s suite of statistical forecasting models,” *Economic Modelling* **25**(4), 772–792.
- KAPETANIOS G, MITCHELL J, PRICE S, AND FAWCETT N (2015), “Generalized density forecast combinations,” *Journal of Econometrics* **188**, 150–165.
- KASTNER G, AND FRÜHWIRTH-SCHNATTER S (2014), “Ancillarity-sufficiency interweaving strategy (ASIS) for boosting MCMC estimation of stochastic volatility models,” *Computational Statistics & Data Analysis* **76**, 408–423.
- KASTNER G, FRÜHWIRTH-SCHNATTER S, AND LOPES HF (2017), “Efficient Bayesian inference for multivariate factor stochastic volatility models,” *Journal of Computational and Graphical Statistics* **26**(4), 905–917.
- KNAUS P, BITTO-NEMLING A, CADONNA A, AND FRÜHWIRTH-SCHNATTER S (2021), “Shrinkage in the Time-Varying Parameter Model Framework Using the R Package shrinkTVP,” *Journal of Statistical Software* **100**(13), 1–32.
- KNOTEK ES, AND ZAMAN S (2023), “Real-time density nowcasts of US inflation: A model combination approach,” *International Journal of Forecasting* **39**(4), 1736–1760.
- KNÜPPEL M (2015), “Evaluating the calibration of multi-step-ahead density forecasts using raw moments,” *Journal of Business & Economic Statistics* **33**(2), 270–281.

## Bibliography

- KOOP G, AND KOROBILIS D (2012), “Forecasting inflation using dynamic model averaging,” *International Economic Review* **53**, 867–886.
- (2013), “Large time-varying parameter VARs,” *Journal of Econometrics* **177**, 185–198.
- LAVINE I, LINDON M, AND WEST M (2021), “Adaptive variable selection for sequential prediction in multivariate dynamic models,” *Bayesian Analysis* **16**, 1059–1083.
- LI L, KANG Y, AND LI F (2023), “Bayesian forecast combination using time-varying features,” *International Journal of Forecasting* **39**(3), 1287–1302.
- LOAIZA-MAYA R, MARTIN GM, AND FRAZIER DT (2021), “Focused Bayesian prediction,” *Journal of Applied Econometrics* **36**, 517–543.
- LOPES HF (2014), “Modern Bayesian Factor Analysis,” in I JELIAZKOV, AND XS YANG (eds.) “Bayesian Inference in the Social Sciences,” 115–153, Hoboken, NJ, USA: John Wiley & Sons, Inc.
- LOPES HF, AND WEST M (2004), “Bayesian model assessment in factor analysis,” *Statistica Sinica* **14**(1), 41–67, publisher: Institute of Statistical Science, Academia Sinica.
- LÓPEZ-SALIDO JD, AND LORIA F (2020), “Inflation at risk,” Finance and Economics Discussion Series 2020-013, Board of Governors of the Federal Reserve System (U.S.).
- MAKALIC E, AND SCHMIDT DF (2015), “A simple sampler for the horseshoe estimator,” *IEEE Signal Processing Letters* **23**(1), 179–182.
- MARTIN GM, FRAZIER DT, MANEESOONTHORN W, LOAIZA-MAYA R, HUBER F, KOOP G, MAHEU J, NIBBERING D, AND PANAGIOTELIS A (2023), “Bayesian forecasting in economics and finance: A modern review,” *International Journal of Forecasting* **40**, 811–839.

## Bibliography

- MCALINN K (2021), “Mixed-frequency Bayesian predictive synthesis for economic now-casting,” *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **70**(5), 1143–1163.
- MCALINN K, AASTVEIT KA, NAKAJIMA J, AND WEST M (2020), “Multivariate Bayesian predictive synthesis in macroeconomic forecasting,” *Journal of the American Statistical Association* **115**, 1092–1110.
- MCALINN K, AND WEST M (2019), “Dynamic Bayesian predictive synthesis in time series forecasting,” *Journal of Econometrics* **210**(1), 155–169.
- MCCAUSLAND WJ, MILLER S, AND PELLETIER D (2011), “Simulation smoothing for state–space models: A computational efficiency analysis,” *Computational Statistics Data Analysis* **55**(1), 199–212.
- MCCRACKEN MW, AND NG S (2021), “FRED-QD: A quarterly database for macroeconomic research,” *Review* **103**(1), 1–44.
- MITCHELL J, AND HALL SG (2005), “Evaluating, comparing and combining density forecasts using the KLIC with an application to the Bank of England and NIESR fan charts of inflation,” *Oxford Bulletin of Economics and Statistics* **67**, 995–1033.
- OELRICH O, VILLANI M, AND ANKARGREN S (2023), “Local prediction pools,” *Journal of Forecasting* .
- ONORANTE L, AND RAFTERY AE (2016), “Dynamic model averaging in large model spaces using dynamic Occam’s window,” *European Economic Review* **81**, 2–14.
- PRADO R, AND WEST M (2010), *Time Series: Modeling, Computation, and Inference*, Boca Raton: Chapman and Hall/CRC, 1st edition edition.
- RAFTERY AE, KÁRNÝ M, AND ETTLER P (2010), “Online prediction under model uncertainty via dynamic model averaging: Application to a cold rolling mill,” *Technometrics* **52**, 52–66.

## Bibliography

- RAGONNEAU TM, AND ZHANG Z (2023), “PDFO: A cross-platform package for Powell’s derivative-free optimization solvers,” [Arxiv.org/abs/2302.13246](https://arxiv.org/abs/2302.13246).
- RAY P, AND BHATTACHARYA A (2018), “Signal adaptive variable selector for the horse-shoe prior,” Technical Report arXiv:1810.09004, arXiv.
- ROBERTS GO, AND ROSENTHAL JS (2009), “Examples of adaptive MCMC,” *Journal of Computational and Graphical Statistics* **18**(2), 349–367.
- ROSSI B (2021), “Forecasting in the presence of instabilities: How we know whether models predict well and how to improve them,” *Journal of Economic Literature* **59**(4), 1135–90.
- ROSSI B, AND SEKHPOSYAN T (2014), “Evaluating predictive densities of US output growth and inflation in a large macroeconomic data set,” *International Journal of Forecasting* **30**(3), 662–682.
- (2019), “Alternative tests for correct specification of conditional predictive densities,” *Journal of Econometrics* **208**(2), 638–657.
- RUBIN DB, GELMAN A, CARLIN JB, STERN HS, DUNSON DB, AND VEHTARI A (2015), *Bayesian Data Analysis*, New York: Chapman and Hall/CRC, 3 edition.
- RUE H, AND HELD L (2005), *Gaussian Markov Random Fields: Theory and Applications*, New York: Chapman and Hall/CRC.
- SMITH JQ (1979), “A generalization of the Bayesian steady forecasting model,” *Journal of the Royal Statistical Society (Series B: Methodological)* **41**, 375–387.
- STOCK JH, AND WATSON MW (2003), “Forecasting output and inflation: The role of asset prices,” *Journal of Economic Literature* **41**(3), 788–829.
- (2007), “Why Has U.S. Inflation Become Harder to Forecast?” *Journal of Money, Credit and Banking* **39**(s1), 3–33.

## Bibliography

- TAKANASHI K, AND MCALINN K (2021), “Predictions with dynamic Bayesian predictive synthesis are exact minimax,” *arXiv:1911.08662 [econ, math, stat]* ArXiv:1911.08662.
- TALLMAN E (2024), *Bayesian Predictive Decision Synthesis: Methodology and Applications*, Ph.D. thesis, Duke University, [stat.duke.edu/alumni/alumni-lists-theses/phd-year](http://stat.duke.edu/alumni/alumni-lists-theses/phd-year).
- TALLMAN E, AND WEST M (2022), “On entropic tilting and predictive conditioning,” *Technical Report, Duke University* [Arxiv.org/abs/2207.10013](https://arxiv.org/abs/2207.10013).
- (2023), “Bayesian predictive decision synthesis,” *Journal of the Royal Statistical Society Series B: Statistical Methodology* **86**, 340–363.
- (2024), “Predictive decision synthesis for portfolios: Betting on better models,” *Technical Report, Duke University, Submitted for publication* .
- TIMMERMANN A (2006a), “Forecast Combinations,” *Handbook of Economic Forecasting*, Elsevier.
- (2006b), “Forecast Combinations,” in G ELLIOTT, CW GRANGER, AND A TIMMERMANN (eds.) “*Handbook of Economic Forecasting*,” volume 1, 135–196, Elsevier.
- VEHTARI A, AND SÄRKKÄ S (2014), “MCMC diagnostics toolbox for Matlab 6.x,” <https://users.aalto.fi/~ave/code/mcmcdiag/>.
- WALLIS KF (2005), “Combining density and interval forecasts: A modest proposal\*,” *Oxford Bulletin of Economics and Statistics* **67**(s1), 983–994.
- WEST M (1992), “Modelling agent forecast distributions,” *Journal of the Royal Statistical Society: Series B (Methodological)* **54**(2), 553–567.
- (2023), “Perspectives on constrained forecasting,” *Bayesian Analysis* [Doi.org/10.1214/23-BA1379](https://doi.org/10.1214/23-BA1379).

## Bibliography

- WEST M, AND CROSSE J (1992), “Modelling probabilistic agent opinion,” *Journal of the Royal Statistical Society: Series B (Methodological)* **54**(1), 285–299.
- WEST M, AND HARRISON PJ (1989), *Bayesian Forecasting and Dynamic Models*, Springer-Verlag, 1 edition.
- (1997), *Bayesian Forecasting and Dynamic Models*, Springer, 2 edition.
- WU JC, AND XIA FD (2016), “Measuring the Macroeconomic Impact of Monetary Policy at the Zero Lower Bound,” *Journal of Money, Credit and Banking* **48**, 253–291.
- YU Y, AND MENG XL (2011), “To center or not to center: That is not the question—Ancillarity—sufficiency interweaving strategy (ASIS) for boosting MCMC efficiency,” *Journal of Computational and Graphical Statistics* **20**(3), 531–570.
- ZHAO ZY, XIE M, AND WEST M (2016), “Dynamic dependence networks: Financial time series forecasting and portfolio decisions (with discussion),” *Applied Stochastic Models in Business and Industry* **32**, 311–339.