Chernis, Tony

Working Paper

Combining large numbers of density predictions with Bayesian predictive synthesis

Bank of Canada Staff Working Paper, No. 2023-45

Provided in Cooperation with:
Bank of Canada, Ottawa

Suggested Citation: Chernis, Tony (2023) : Combining large numbers of density predictions with Bayesian predictive synthesis, Bank of Canada Staff Working Paper, No. 2023-45, Bank of Canada, Ottawa,
https://doi.org/10.34989/swp-2023-45

This Version is available at:
https://hdl.handle.net/10419/297430

BANK OF CANADA
BANQUE DU CANADA

Staff Working Paper/Document de travail du personnel—2023-45

Last updated: August 21, 2023

Combining Large Numbers of Density Predictions with Bayesian Predictive Synthesis

by Tony Chernis

Canadian Economic Analysis Department, Bank of Canada and
University of Strathclyde
tchernis@bankofcanada.ca

DOI: https://doi.org/10.34989/swp-2023-45 | **ISSN** 1701-9397

Acknowledgements

I am extremely grateful to Gary Koop and Stuart McIntyre for their guidance, comments and feedback through the development of this paper. I would also like to thank Rodrigo Sekkel, Luis Uzeda, and seminar participants at the University of Strathclyde, Bank of Canada, Orebro Workshop on Financial Econometrics, and the Conference on Real-Time Data at the Cleveland Federal Reserve Bank whose comments have greatly improved this paper. I am also thankful for the excellent research assistance of Tasha Reader. Any remaining errors are the responsibility of the authors. The views expressed in this paper are solely those of the authors. No responsibility for them should be attributed to the Bank.

Abstract

Bayesian predictive synthesis is a flexible method of combining density predictions. The flexibility comes from the ability to choose an arbitrary synthesis function to combine predictions. I study the choice of synthesis function when combining large numbers of predictions—a common occurrence in macroeconomics. Estimating combination weights with many predictions is difficult, so I consider shrinkage priors and factor modelling techniques to address this problem. The dense weights of factor modelling provide an interesting contrast with the sparse weights implied by shrinkage priors. I find that the sparse weights of shrinkage priors perform well across exercises.

Topics: Econometric and statistical methods
JEL codes: C11, C52, C53, E37

Résumé

La synthèse de prévisions bayésienne est une méthode servant à combiner des prédictions de densités. Comme elle permet de sélectionner une fonction de synthèse arbitraire pour ce faire, elle offre une certaine souplesse. Je me penche sur la sélection de cette fonction dans un contexte où il faut combiner de nombreuses prévisions, ce qui est courant en macroéconomie. Il est difficile d'estimer la pondération des combinaisons pour un grand nombre de prévisions. J'évalue donc des mesures de rétrécissement a priori et des techniques de modélisation factorielle pour surmonter cette difficulté. Les fortes pondérations issues de la modélisation factorielle offrent un contraste intéressant avec les faibles pondérations que laissent sous-entendre les mesures de rétrécissement a priori. Je constate que ces faibles pondérations donnent de bons résultats pour l'ensemble des exercices d'établissement de prévisions.

Sujets : Méthodes économétriques et statistiques
Codes JEL : C11, C52, C53, E37

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.

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), 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.*, 2021). Bayesian approaches can also face difficulties in large dimensions. For example, Bayesian Model Averaging requires calculation of the marginal likeli-

hood 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 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.[1] 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.[2] Furthermore, Knotek and Zaman (2022) 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.* (2018).

The remainder of the paper proceeds as follows. Section 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

[1] For example, Wallis (2005); Hall and Mitchell (2007); Mitchell and Hall (2005); Bache *et al.* (2009).
[2] Del Negro *et al.* (2016); Billio *et al.* (2013); Aastveit *et al.* (2016); McAlinn and West (2019).

provide a brief overview of global-local shrinkage priors and Bayesian factor models. Section 3 details the prediction exercises, while Section 4 discusses the results. Section 5 concludes.

2 Econometric Framework

In this section, I begin by describing BPS, followed by a discussion of the synthesis functions. Appendix A 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.

2.1 Bayesian Predictive Synthesis

Bayesian Predictive Synthesis is a method for combining predictive densities.[3] 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:

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

where $x_t = x_{t,1:J} = (1, x_{t,1}, \ldots, x_{t,J})'$, 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. $\alpha(y_t|x_t)$ is an arbitrary synthesis

[3] 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.

function used to combine the expert densities, while Φ_t are the 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 1 in hand, I can write out a Gibbs Sampler with two blocks:

1. Estimate the synthesis function $\alpha(y_t|x_t)$ 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):

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

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, β_t are combination weights that vary over time following a random walk with variance θ, and ϵ_t is an error term with time varying volatility σ_t^2.

The first step in BPS is to estimate equation 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 $p(x_{1:t}|\Phi_{1:t}, y_{1:t}, \mathcal{H}_{1:t})$ conditional on the values of the synthesis function parameters (Φ_t). These x_t are conditionally independent over time with the following conditionals:

$$p(x_t|\Phi_t, y_t, \mathcal{H}_t) \propto N(y_t|X_t'\beta_t, \epsilon_t) \prod_{j=1:J} h_{tj}(x_{tj}) \quad \text{with} \quad x_t = (1, x_{t1}, \ldots, x_{tJ})' \tag{3}$$

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 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 κ is a global shrinkage parameter and ψ_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 is 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):

$$\begin{aligned} y_t &= x_t\beta + x_t Diag(\sqrt{\theta_1}, ..., \sqrt{\theta_d})\tilde{\beta}_t + \epsilon_t, \qquad \epsilon_t \sim \mathcal{N}(0, \sigma_t^2) \\ \tilde{\beta}_t &= \tilde{\beta}_{t-1} + \tilde{u}_t, \qquad \tilde{u}_t \sim \mathcal{N}_J(0, I_J) \end{aligned} \tag{4}$$

The non-centered parameterization allows shrinkage on θ, which is the variance of β_t, and β, 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 (2017), Kastner *et al.* (2017)). The model in equation 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.

2.3 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. 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 5:

$$y_t = F_t^{'}\gamma_t + \epsilon_t \qquad \gamma_t = \gamma_{t-1} + u_t \qquad x_t = \Lambda f_t + \nu_t \tag{5}$$

$$\epsilon_t \sim \mathcal{N}(0, \sigma_t^2) \qquad u_t \sim \mathcal{N}(0, \theta) \qquad \nu_t \sim \mathcal{N}(0, R) \tag{6}$$

where $F_t = (1, f_t)$ and f_t is a $k \times 1$ vector of factors, Λ is a $J \times k$ vector of loadings, and R is a diagonal covariance matrix with elements $\sigma_{\nu J}^2$. Additionally, γ_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 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.

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.[4] 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.[5] 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-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.

[4]Examples are Aastveit *et al.* (2011), Chernis and Webley (2022), and Knotek and Zaman (2022).
[5]For example, Diebold *et al.* (2021) and Conflitti *et al.* (2015) consider the European SPF.

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 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.

In a nowcasting exercise, the timing of the forecast cycle can be quite important. Figure 2 illustrates the timing of releases throughout the six-month forecast cycle 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.

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).

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.* (2021) 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.[6] 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.* (2021), 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.

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 results for the nowcasting exercise are reported in table 1, and table 3 shows results for the survey forecast application. The tables also include as a benchmark the dynamic linear model (DLM) in equation 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

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

calibrated using the test from Knüppel (2015).

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 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 excercise 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 5 shows results for the nowcasting application and figure 6 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 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 application and the longer prediction horizons in the nowcasting application. Inspecting the forecast densities over time can help explain these results.

Figure 7 shows predictive densities for the nowcasting application[7] and figure 9 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

[7]COVID-19 is excluded because the volatility in GDP growth makes the chart illegible.

region by the two factor model. Similar charts for the SPF application are shown in figures 8 and 10.

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.[8] 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 systematically better forecasts during normal times. This can be seen in figure 10 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 6 shows that during the Euro Area crisis, the factor model improves over the triple gamma approach. This can be seen in figure 10 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 5, where the blue line approaches the y-axis at the end of sample, signifying that on average the two 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

[8]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.

slightly exceed the average performance of the shrinkage priors. Figure 8 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 8, 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 10). 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.

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 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 11 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 12). In contrast, the upper panel of figure 12 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.

4.3 Examining the Combination Weights

It is instructive to examine the weights in figure 12 to gain some understanding 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.[9] 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.

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.[10] 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

[9] 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.

[10] 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.

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.

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 struc-

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

References

Aastveit KA, Cross JL, and Dijk HKV (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, Mitchell J, Ravazzolo F, and Van Dijk HK (2018), "The evolution of forecast density combinations in economics," Technical Report 18-069/III, Tinbergen Institute.

Aastveit KA, Ravazzolo F, and Van Dijk HK (2016), "Combined density nowcasting in an uncertain economic environment," *Journal of Business & Economic Statistics* **0**(0), 1–15.

Bache IW, Mitchell J, Ravazzolo F, and Vahey SP (2009), "Macro modelling with many models," Technical Report 2009/15, Norges Bank.

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.

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.

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), 20, number: 2 Publisher: Multidisciplinary Digital Publishing Institute.

Carvalho CM, Polson NG, and Scott JG (2010), "The horseshoe estimator for sparse signals," *Biometrika* **97**(2), 465–480, publisher: [Oxford University Press, Biometrika Trust].

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.

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 2022-12, Bank of Canada, publication Title: Discussion Papers.

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**(4), 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.

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, Shin M, and Zhang B (2021), "On the aggregation of probability assessments: Regularized mixtures of predictive densities for Eurozone inflation and real interest rates," Technical Report 21-06, Federal Reserve Bank of Philadelphia, publication Title: Working Papers.

ECB (2019), "Results of the third special questionnaire for participants in the ECB Survey of Professional Forecasters," Https://www.ecb.europa.eu/stats/ecb_surveys/survey_of_professional_forecasters/html/ecb.spf201902_speci

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.

García JA (2003), "An introduction to the ECB's survey of professional forecasters," Occasional Paper Series 8, European Central Bank.

Geweke J, and Amisano G (2011), "Optimal prediction pools," *Journal of Econometrics* **164**(1), 130–141.

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, publisher: Taylor & Francis _eprint: https://doi.org/10.1198/016214506000001437.

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.

Hörmann W, and Leydold J (2014), "Generating generalized inverse Gaussian random variates," *Statistics and Computing* **24**(4), 547–557.

Jore AS, Mitchell J, and Vahey SP (2010), "Combining forecast densities from VARs with uncertain instabilities," *Journal of Applied Econometrics* **25**(4), 621–634.

Kastner G, and Frühwirth-Schnatter S (2017), "Ancillarity-sufficiency interweaving strategy (ASIS) for boosting MCMC estimation of Stochastic volatility models," Paper, arXiv.org.

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**, 1–32.

Knotek ES, and Zaman S (2022), "Real-time density nowcasts of US inflation: A model combination approach," *International Journal of Forecasting* .

Knüppel M (2015), "Evaluating the calibration of multi-step-ahead density forecasts using raw moments," *Journal of Business & Economic Statistics* **33**(2), 270–281.

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.

McAlinn K (2021), "Mixed-frequency Bayesian predictive synthesis for economic nowcasting," *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **70**(5), 1143–1163, https://onlinelibrary.wiley.com/doi/pdf/10.1111/rssc.12500.

McAlinn K, Aastveit KA, Nakajima J, and West M (2020), "Multivariate Bayesian predictive synthesis in macroeconomic forecasting," *Journal of the American Statistical Association* **115**(531), 1092–1110, publisher: Taylor & Francis _eprint: https://doi.org/10.1080/01621459.2019.1660171.

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.

Mitchell J, and Hall S (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**(s1), 995–1033.

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.

Ray P, and Bhattacharya A (2018), "Signal adaptive variable selector for the horseshoe prior," Technical Report arXiv:1810.09004, arXiv, arXiv:1810.09004 [stat] type: article.

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.

Takanashi K, and McAlinn K (2021), "Predictions with dynamic Bayesian predictive synthesis are exact minimax," *arXiv:1911.08662 [econ, math, stat]* ArXiv: 1911.08662.

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.

West M, and Crosse J (1992), "Modelling probabilistic agent opinion," *Journal of the Royal Statistical Society: Series B (Methodological)* **54**(1), 285–299.

Yu Y, and Meng XL (2011), "To center or not to center: That is not the question—An ancillarity—sufficiency interweaving strategy (ASIS) for boosting MCMC efficiency," *Journal of Computational and Graphical Statistics* **20**(3), 531–570, publisher: [American Statistical Association, Taylor & Francis, Ltd., Institute of Mathematical Statistics, Interface Foundation of America].

6 Figures

Figure 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

Figure 2: Overview of Forecast Cycle

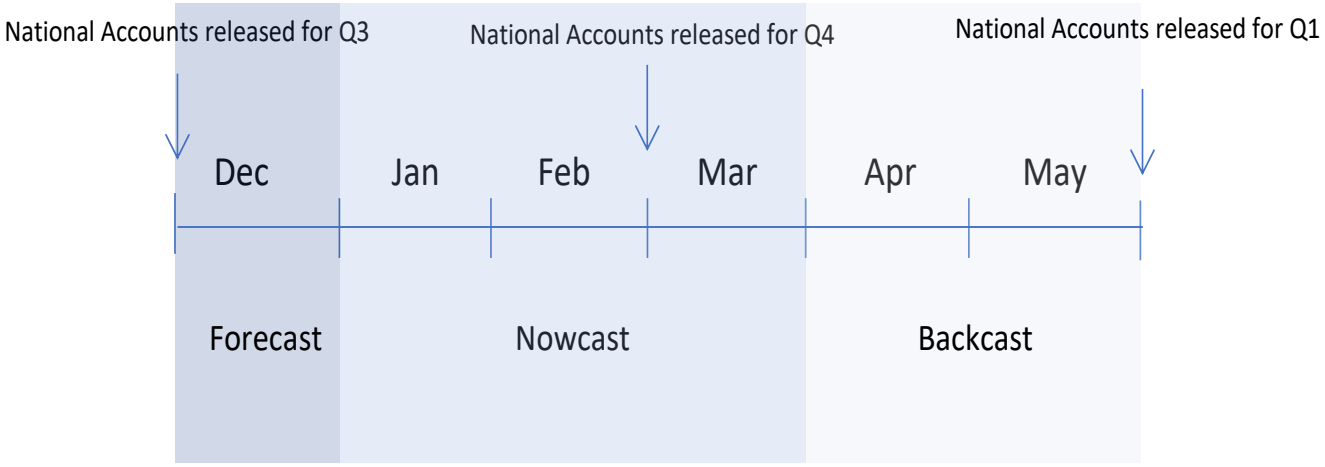
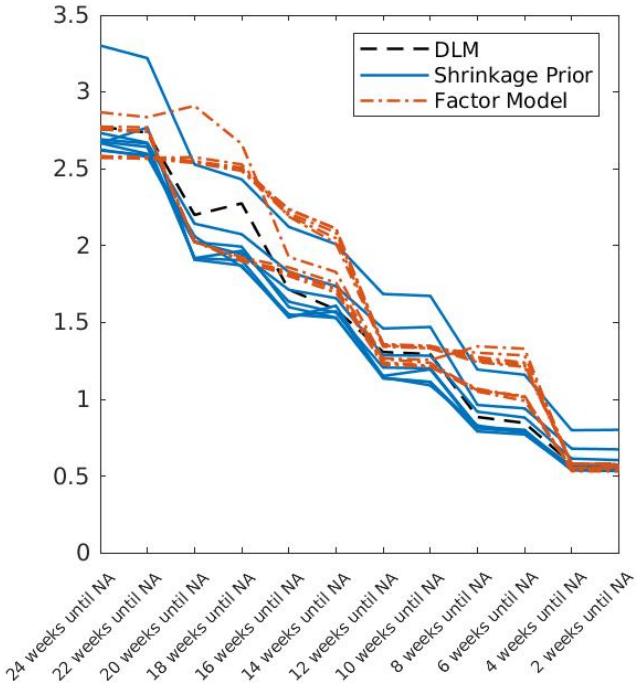
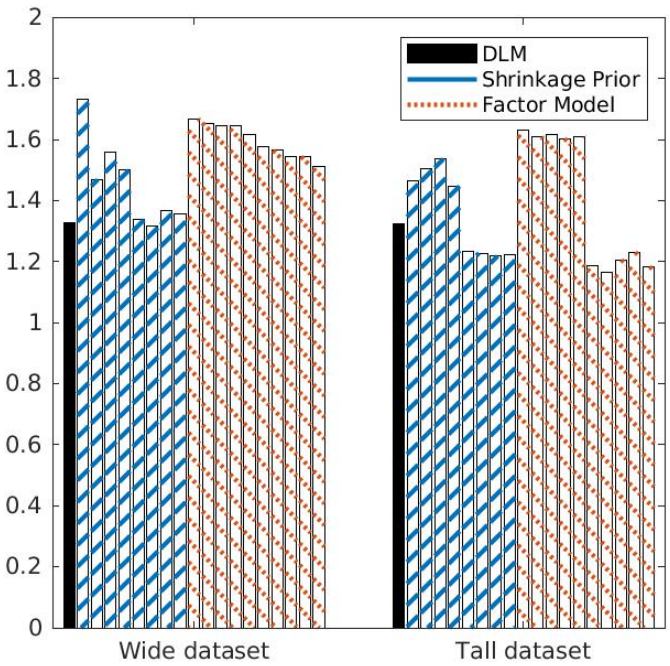


Figure 3: Comparison of Shrinkage Priors and Factor Models

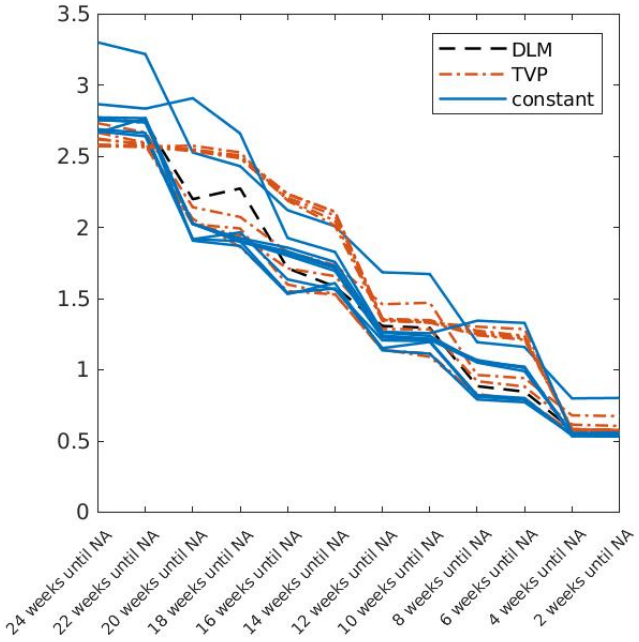


(a) CRPS in Nowcasting Application

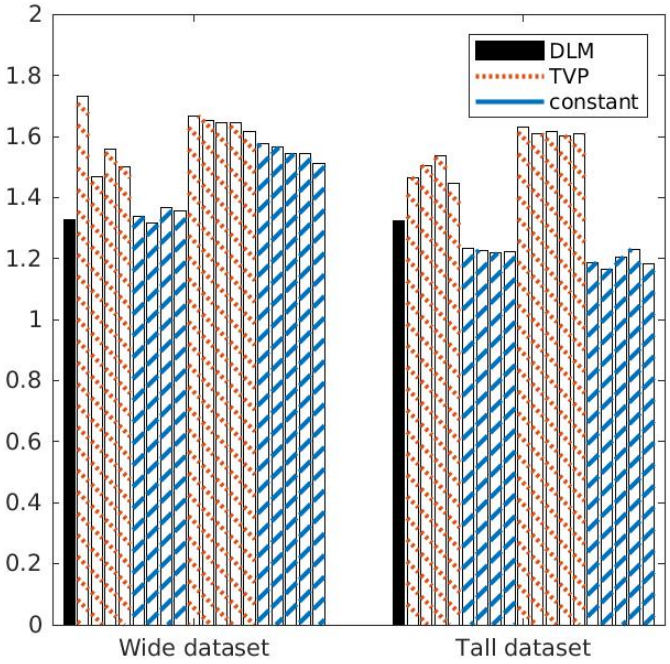


(b) CRPS in SPF Application

Figure 4: Comparison of Time-varying and Constant Parameter Specification



(a) CRPS in Nowcasting Exercise



(b) CRPS for SPF forecasts

Table 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 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 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.

Figure 5: Nowcasting Cumulative CRPS Difference: Factor Model 2 (constant) - Triple Gamma (constant)

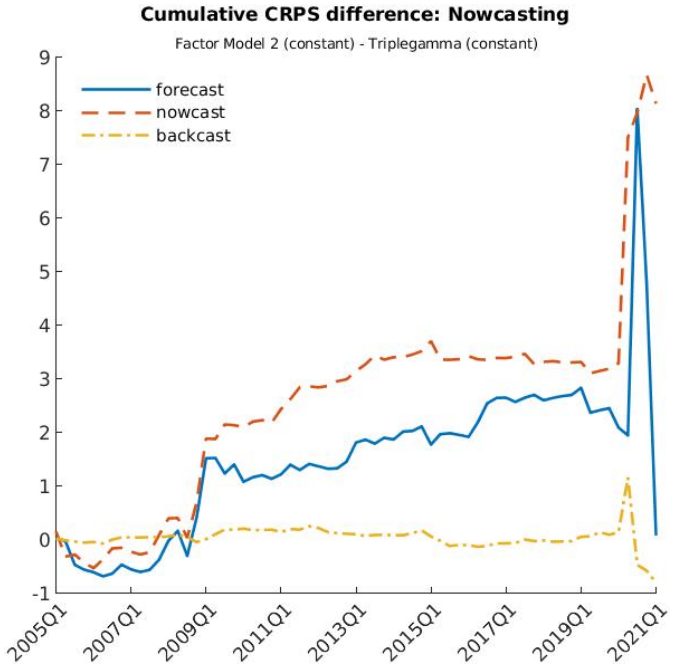


Figure 6: SPF Cumulative CRPS Difference: Factor Model 2 (constant) - Triple Gamma (constant)

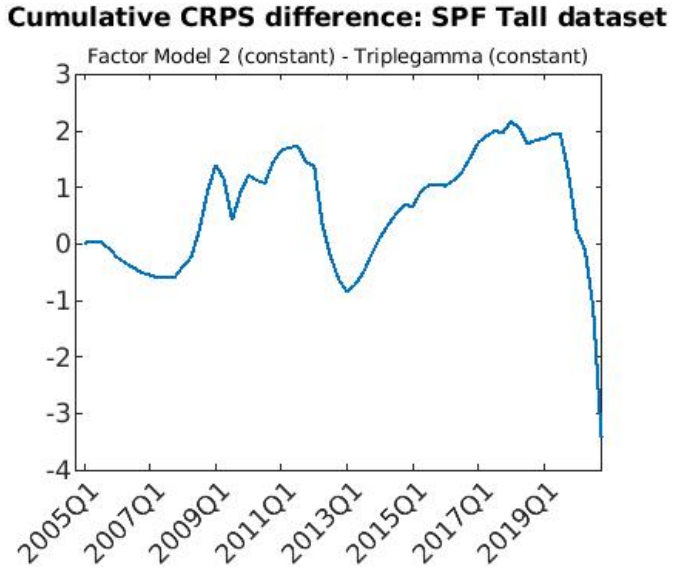


Figure 7: Nowcasting Application Predictive Densities

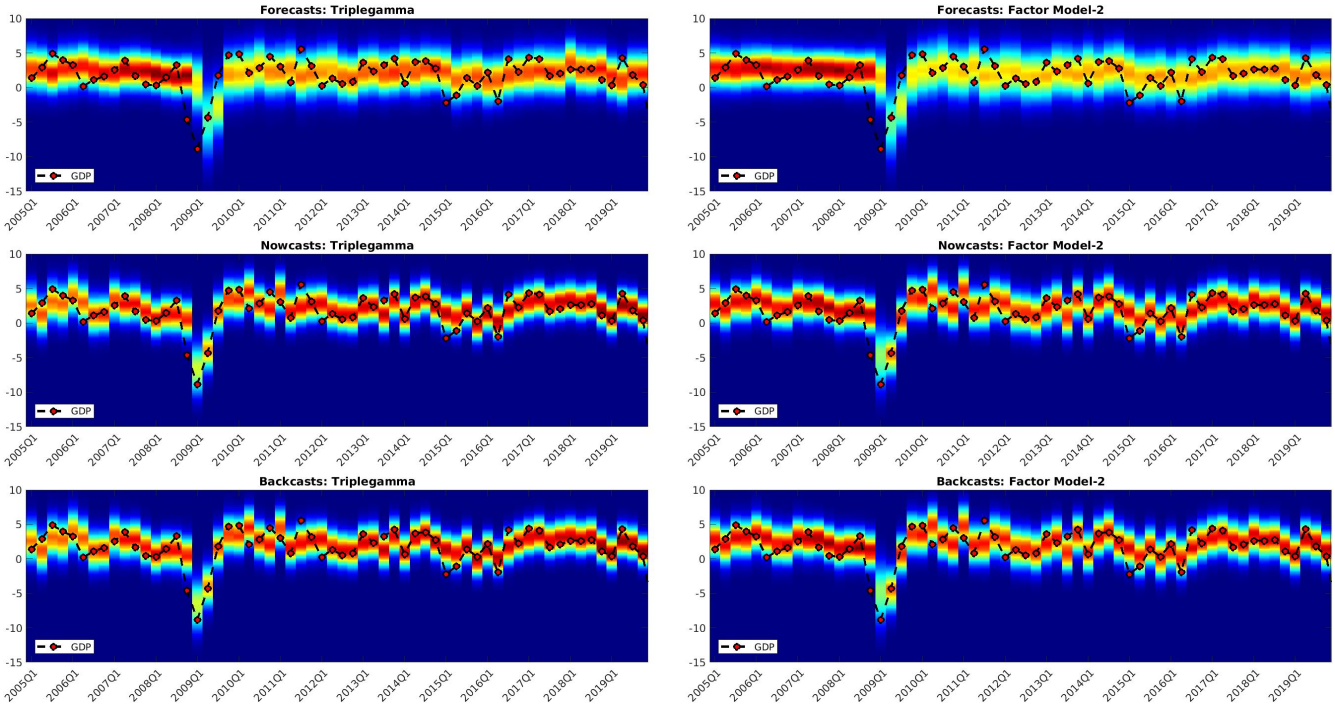


Figure 8: SPF Application Predictive Densities

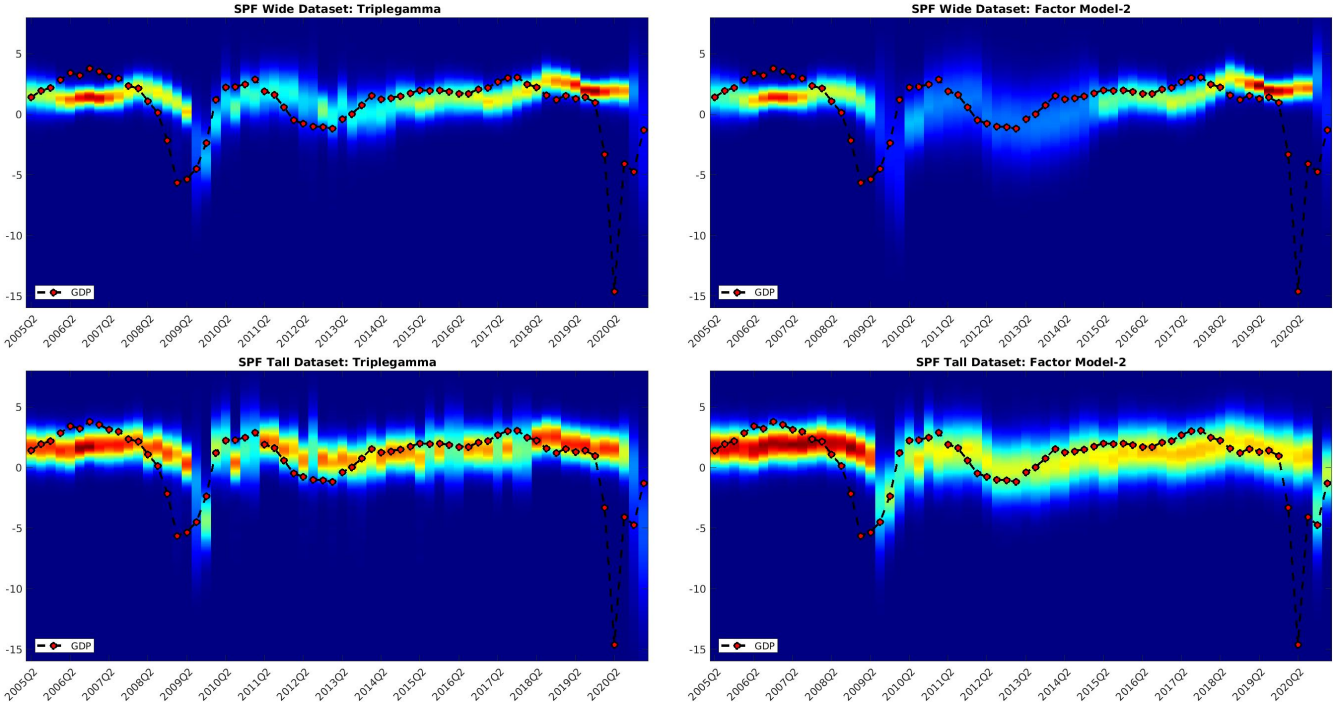
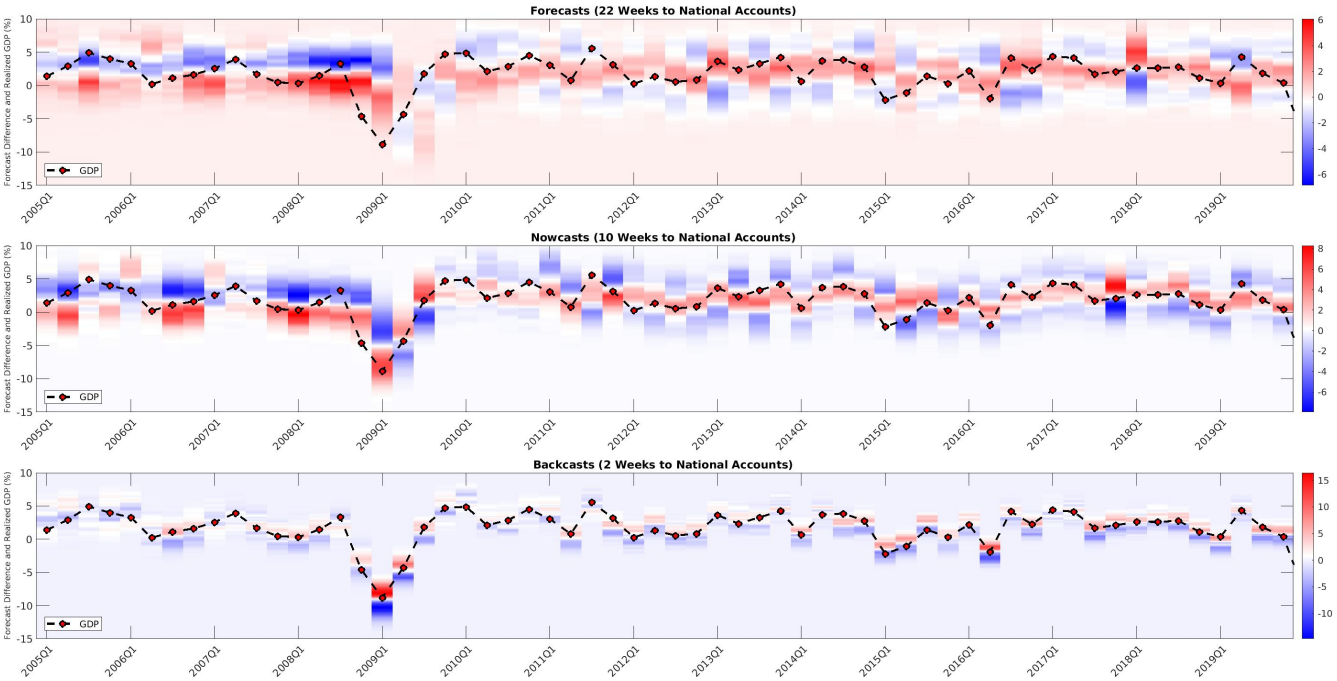
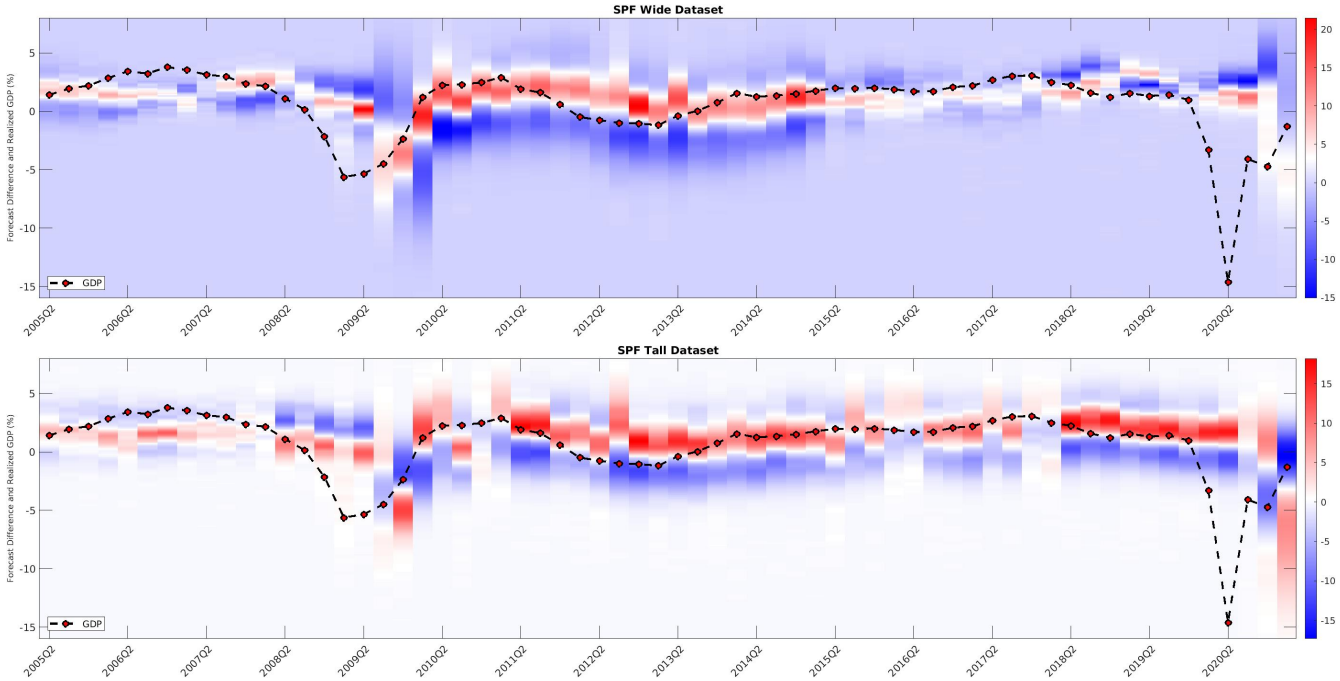


Figure 9: Nowcasting Application Difference Between Triplegamma and Factor Model-2 in Predictive Densities with GDP Realizations



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 10: **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.

Figure 11: Time-varying Predictive of Mean of BPS Intercepts for SPF Tall Dataset

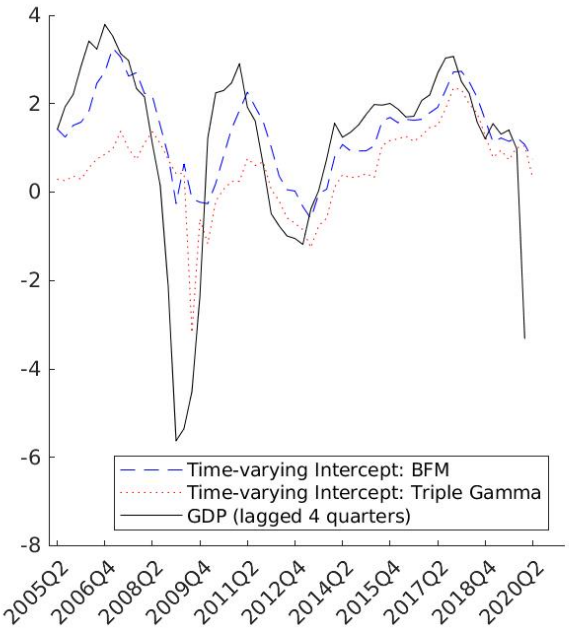


Figure 12: Sequentially Estimated Mean Combination Weights

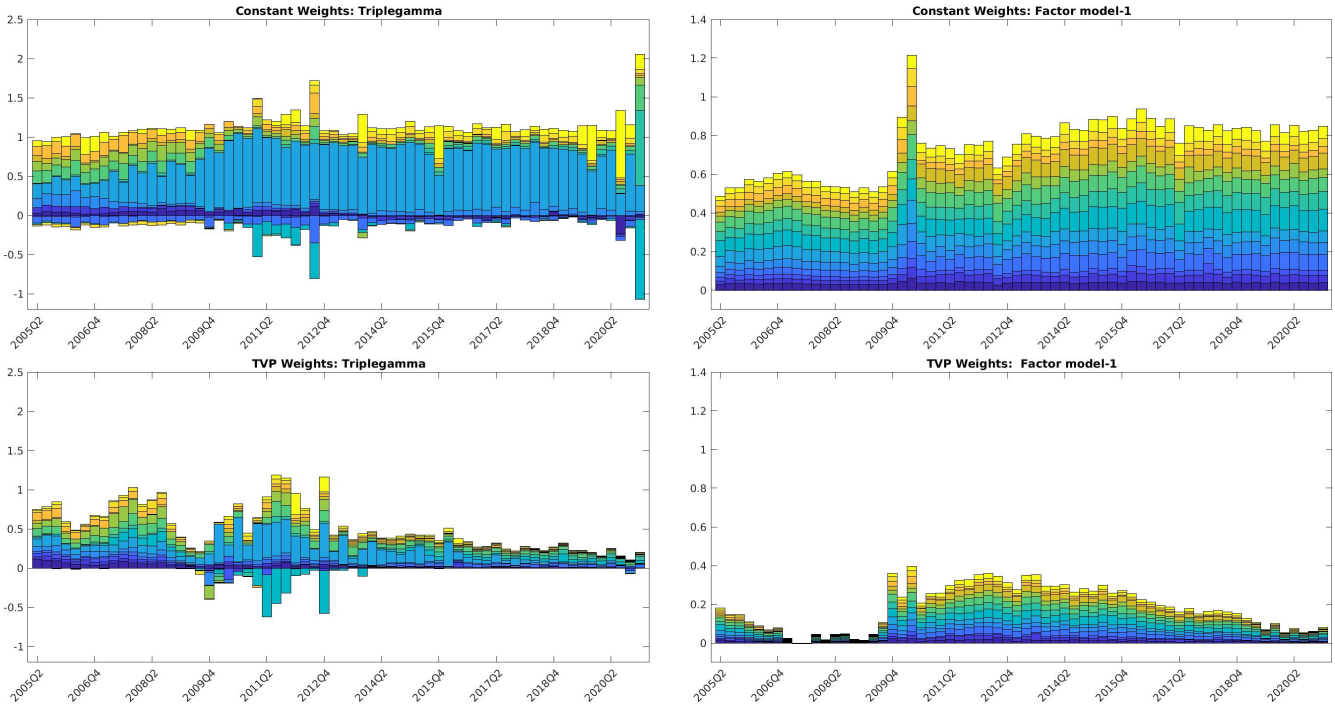
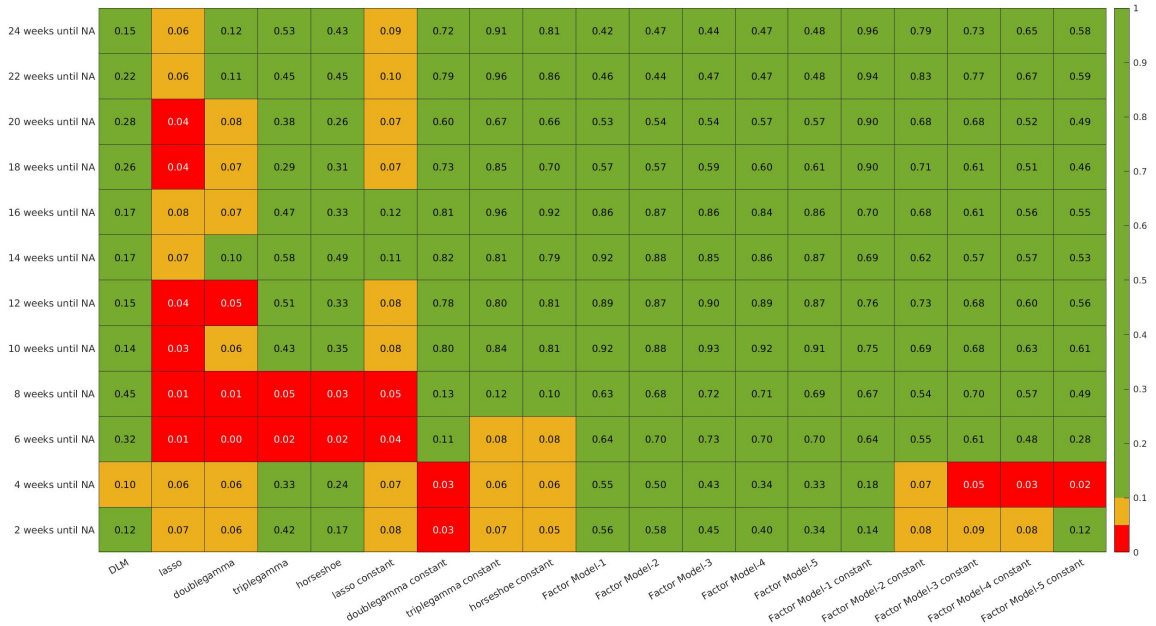
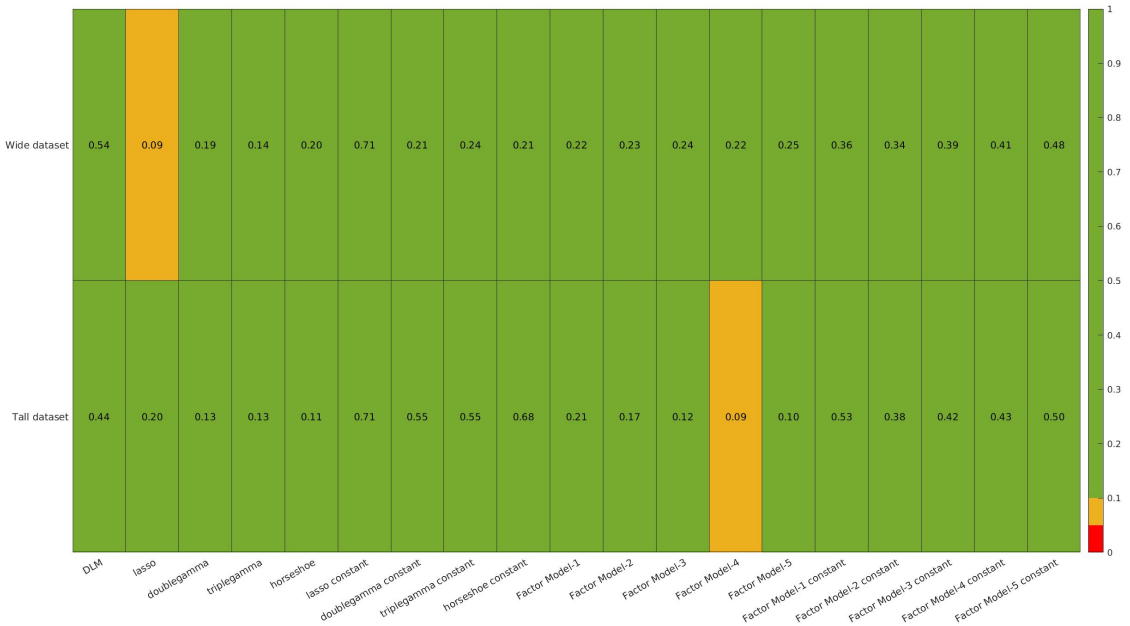


Figure 13: Knüpple Test for Probabilistic Calibration: Nowcasting Application



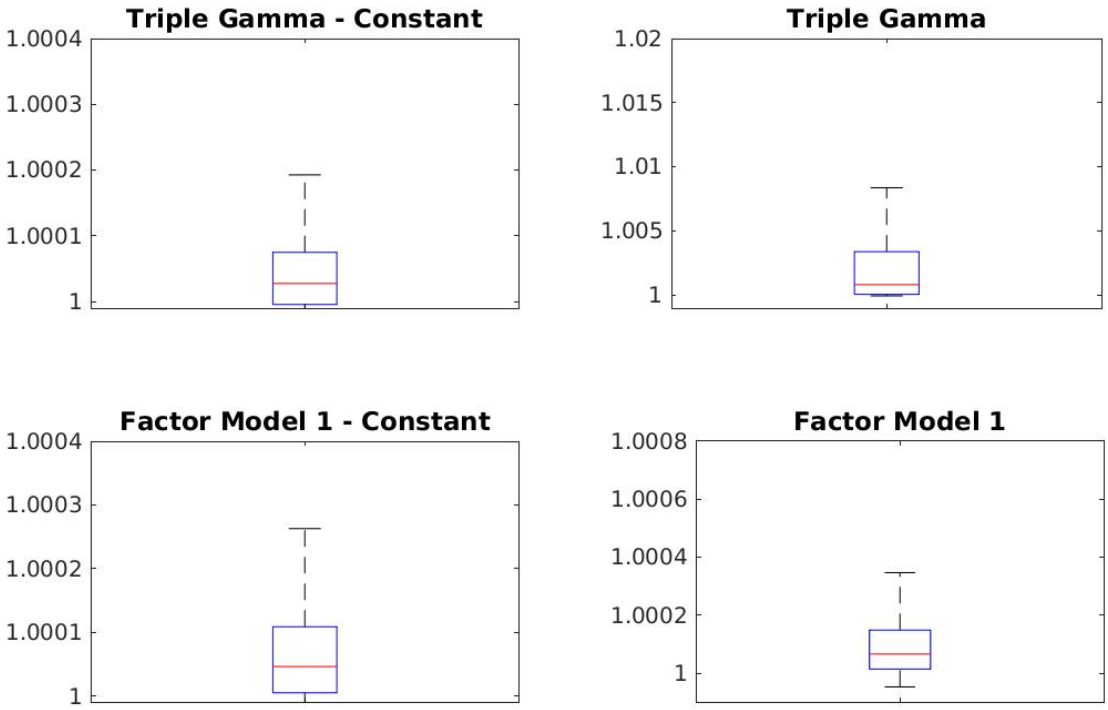
Notes: Results from the Knüpple test for probabilistic calibration. Null hypothesis is for calibration and values in the table correspond to p-values. Red shading corresponds to rejection of calibration at 5 percent level and yellow at 10 percent level.

Figure 14: 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.

Figure 15: Box Plots of Potential Scale Reduction Factors



A Technical Appendix

A.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 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

$$y_t = x_t\beta_t + \epsilon_t \qquad \beta_t = \beta_{t-1} + u_t \qquad\qquad \epsilon_t \sim \mathcal{N}(0, \sigma_t^2) \qquad u_t \sim \mathcal{N}(0, Q) \tag{7}$$

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 (7) can be rewritten equivalently in the non-centered parametrization as

$$\begin{aligned} y_t &= x_t\beta + x_t Diag(\sqrt{\theta_1}, ..., \sqrt{\theta_d})\tilde{\beta}_t + \epsilon_t, \qquad \epsilon_t \sim \mathcal{N}(0, \sigma_t^2) \\ \tilde{\beta}_t &= \tilde{\beta}_{t-1} + \tilde{u}_t, \qquad \tilde{u}_t \sim \mathcal{N}_d(0, I_d) \end{aligned} \tag{8}$$

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,

$$h_t | h_{t-1}, \sigma_\eta^2 \sim \mathcal{N}\left(h_{t-1}, \sigma_\eta^2\right), \tag{9}$$

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

A.2.1 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 β_j, for $j = 1, \ldots, d$. Note that, in the case of the standard deviations, this can equivalently be seen as a triple gamma prior on the innovation variances θ_j, for $j = 1, \ldots, d$. In the constant parameterizations, I place an NGG prior on the β_j using the centered parameterization:

$$\sqrt{\theta}_j | \xi_j^2 \sim \mathcal{N}(0, \xi_j^2), \qquad \xi_j^2 | a^\xi, \kappa_j^2 \sim \mathcal{G}(a^\xi, \frac{a^\xi \kappa_j^2}{2}), \quad \kappa_j^2 | c^\xi, \kappa_B^2 \sim \mathcal{G}(c^\xi, \frac{c^\xi}{\kappa_B^2}) \tag{10}$$

$$\beta_j | \tau_j^2 \sim \mathcal{N}(0, \tau_j^2), \qquad \tau_j^2 | a^\tau, \lambda_j^2 \sim \mathcal{N}(a^\tau, \frac{a^\tau \lambda_j^2}{2}) \quad \lambda_j^2 | c^\tau, \lambda_B^2 \sim \mathcal{N}(c^\tau, \frac{c^\tau}{\lambda_B^2}). \tag{11}$$

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

$$\sqrt{\theta}_j | \xi_j^2 \sim \mathcal{N}(0, \xi_j^2), \qquad \xi_j^2 | a^\xi, \kappa_B^2 \sim \mathcal{G}(a^\xi, \frac{a^\xi \kappa_B^2}{2}), \tag{12}$$

$$\beta_j | \tau_j^2 \sim \mathcal{N}(0, \tau_j^2), \qquad \tau_j^2 | a^\tau, \lambda_B^2 \sim \mathcal{G}(a^\tau, \frac{a^\tau \lambda_B^2}{2}). \tag{13}$$

The parameters a^ξ, a^τ, c^ξ, c^τ, κ_B^2, and λ_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:

$$\frac{\kappa_B^2}{2}|a^\xi, c^\xi \sim F(2a^\xi, 2c^\xi), \qquad \frac{\lambda_B^2}{2}|a^\tau, c^\tau \sim F(2a^\tau, 2c^\tau). \tag{14}$$

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

$$2a^\xi \sim \mathcal{B}\left(\alpha_{a^\xi}, \beta_{a^\xi}\right), \qquad 2c^\xi \sim \mathcal{B}\left(\alpha_{c^\xi}, \beta_{c^\xi}\right), \tag{15}$$

$$2a^\tau \sim \mathcal{B}\left(\alpha_{a^\tau}, \beta_{a^\tau}\right), \qquad 2c^\tau \sim \mathcal{B}\left(\alpha_{c^\tau}, \beta_{c^\tau}\right). \tag{16}$$

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 parameters follow independent gamma distributions:

$$\kappa_B^2 \sim \mathcal{G}(d_1, d_2), \qquad \lambda_B^2 \sim \mathcal{G}(e_1, e_2). \tag{17}$$

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:

$$a^\xi \sim \mathcal{G}(\alpha_{a^\xi}, \alpha_{a^\xi}\beta_{a^\xi}), \qquad a^\tau \sim \mathcal{G}(\alpha_{a^\tau}, \alpha_{a^\tau}\beta_{a^\tau}), \tag{18}$$

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 α_{a^ξ} and α_{a^τ} 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 σ^2 follows a gamma distribution; that is,

$$\sigma^2|C_0 \sim \mathcal{G}^{-1}(c_0, C_0), \qquad C_0 \sim \mathcal{G}(c_0 + g_0, (G_0 + \sigma^{-2})^{-1}), \tag{19}$$

with hyperparameters c_0, g_0, and G_0.

In the case of stochastic volatility, the priors on the parameters σ_η^2 in Equation 9 are,

$$\sigma_\eta^2 \sim \mathcal{G}^{-1}(\nu, S_h), \qquad h_0 \sim \mathcal{N}(a_0, b_0) \tag{20}$$

with hyperparameters ν, S_h, a_0 and b_0.

A.2.2 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 $\tilde{\beta} = (\tilde{\beta}_0, \ldots, \tilde{\beta}_T)$ 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 $\beta_1, \ldots, \beta_d$, and $\sqrt{\theta_1}, \ldots, \sqrt{\theta_d}$ in the non-centered parametrization from a multivariate normal distribution. Otherwise, sample $\beta_1, \ldots, \beta_d$, in the centered parameterization from a multivariate normal distribution.
3. If in TVP specification, perform an ancillarity-sufficiency interweaving step and redraw each $\beta_1, \ldots, \beta_d$ from a normal distribution and each $\theta_1, \ldots, \theta_d$ 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 $\xi_1^2, \ldots \xi_d^2$ and $\tau_1^2, \ldots \tau_d^2$ 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 σ^2 from an inverse gamma distribution in the homoscedastic case or, in the SV case, sample the volatility of the volatility σ_η^2 and the log-volatilities $h = (h_0, \ldots, h_T)$.

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 $t = 0, \ldots, T$ 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 $\tilde{\beta} = (\tilde{\beta}_0, \ldots, \tilde{\beta}_T)$, 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.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.[11]

$$y_t = F_t^{'}\gamma_t + \epsilon_t \qquad \gamma_t = \gamma_{t-1} + u_t \qquad x_t = \Lambda f_t + \nu_t \tag{21}$$

$$\epsilon_t \sim \mathcal{N}(0, \sigma_t^2) \qquad u_t \sim \mathcal{N}(0, \theta) \qquad \nu_t \sim \mathcal{N}(0, R) \tag{22}$$

where f_t is a $k \times 1$ vector of factors, $F_t = (1, f_t)$, γ_t is $k+1$ vector of coefficients, Λ 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 Λ, R, σ_t^2, and θ. 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 $i = 1, \ldots, k$. 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 γ_t are $\gamma_0 \sim \mathcal{N}(0, P0)$, where $P0 \sim \mathcal{IG}(\nu_P, (\nu_P - 1) \times c_P)$.

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

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 Λ for $i = 1, \ldots, k$ $\Lambda_i \sim \mathcal{N}(m_i, C_i)1(\Lambda_{ii} > 0)$ where $m_i = C_i(C_0^{-1}\mu_0 1_i + \sigma_{\nu i}^2 F_i x_i)$ and $C_i^{-1} = C_0^{-1} I_i + \sigma_{\nu i}^2 F_i' F_i$.

[11] The implementation in the paper includes an intercept. For ease of exposition, it has been omitted in the following section.

3. Sample Λ for $i = k+1, \ldots, J$ $\Lambda_i \sim \mathcal{N}(m_i, C_i)1(\Lambda_{ii} > 0)$ where $m_i = C_i(C_0^{-1}\mu_0 1_k + \sigma_{\nu i}^2 F' x_i)$ 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 $\gamma_1, \ldots, \gamma_d$, jointly from a multivariate normal distribution using the precision sampler of Chan and Jeliazkov (2009). Otherwise, sample $\gamma = (\gamma_0, \ldots, \gamma_T)$ from a multivariate normal distribution.

6. Sample the error variance σ^2 from an inverse gamma distribution in the homoscedastic case or, in the SV case, sample the volatility of the volatility σ^2 and the log-volatilities $h = (h_0, \ldots, h_T)$.

A.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 Φ 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:

$$p(x_t|\Phi_t, y_t, \mathcal{H}_t) \propto N(y_t|X_t'\beta_t, \epsilon_t) \prod_{j=1:J} h_{tj}(x_{tj}) \quad \text{with} \quad X_t = (1, x_t 1, \ldots, x_{tJ})' \tag{23}$$

If the agents provide normal forecast densities, then 23 yields a multivariate normal distribution for x_t. The posterior distribution for each x_t is:

$$p(x_t|\Phi_t, y_t, \mathcal{H}_t) = \mathcal{N}(h_t + b_t c_t, H_t - b_t b_t' g_t) \tag{24}$$

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 24 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 24 in the MH step. The model is straightforward to re-parameterize with the following steps:

$$y_t = x_t'\gamma_t + \epsilon_t \qquad x_t = \Lambda f_t + \nu_t \tag{25}$$

$$f_t = (\Lambda'\Lambda)^{-1}\Lambda' x_t - (\Lambda'\Lambda)^{-1}\Lambda'\nu_t \quad \text{where,} \quad \Omega = (\Lambda'\Lambda)^{-1}\Lambda' \tag{26}$$

$$y_t = x_t'\Omega'\gamma_t - \nu_t'\Omega'\gamma_t + \epsilon_t \quad \rightarrow \quad y_t = x_t'\gamma_t^* + \epsilon_t^* \tag{27}$$

$$\text{where,} \quad \epsilon_t^* = -\nu_t'\Omega'\gamma_t + \epsilon_t \quad \text{and} \quad \gamma_t^* = \Omega'\gamma_t \tag{28}$$

Now that the model has been re-parameterized, I can use the equation 24 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.

B 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 13 and 14 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.

C 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 15. These results show reasonable convergence of the MCMC algorithms.