Combining Large Numbers of Density Predictions with Bayesian Predictive Synthesis *

Tony Chernis‡

Bank of Canada and University of Strathclyde

June 22, 2022

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 choice of synthesis 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 as ways to address this problem. These techniques provide an interesting contrast between the sparse weights implied by shrinkage priors and dense weights of factor modelling techniques. I find that the sparse weights of shrinkage priors perform well across exercises.

JEL classification: C11, C52, C53, E37
Topics: Density Prediction, Forecast combination, Econometric and statistical methods

*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, Florian Huber, James Mitchell, and seminar participants at the University of Strathclyde and Bank of Canada 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. Bank of Canada Staff Working Papers are completed staff research studies on a wide variety of subjects relevant to central bank policy, produced independently from the Bank's Governing Council. The views expressed in this paper are solely those of the authors and may differ from official Bank of Canada views. No responsibility for them should be attributed to the Bank.

†200 Granville Street, Suite 2160, Vancouver, BC, Canada, V6C 1S4

‡Email: tchernis@bankofcanada.ca Phone: 1-604-643-6231

1 Introduction

Policymakers and academics are interested in combinations of density forecasts. Interest in combining density predictions comes from the recognition that individual models often provide a partial understanding of the economy due to different underlying datasets or modelling assumptions. As a result, practitioners consult a wide variety of models and experts to form the basis of their decision-making (Coletti and Murchison, 2002). Furthermore, density forecasts and their combinations are useful because they fully characterize the uncertainty around a prediction. A density prediction can be very useful to decisionmakers (Chernis and Webley, 2022) since it can show the uncertainty around a prediction, the balance of risks, or the severity of tail risks. However, decision makers often consider large sets of models, commonly seen in nowcasting applications or expert surveys such as the Survey of Professional Forecasters. Combining these large numbers of forecasts can be difficult and requires specialized techniques. A common solution to the problem of processing large amounts of data in economics is shrinkage and factor modelling techniques.

In this paper, I compare global-local shrinkage priors and factor models when combining forecasts within the framework of Bayesian Predictive Synthesis (BPS). Global-local shrinkage priors and factor models naturally lend themselves to large dimensional problems and BPS is an intuitive approach to combining densities. BPS frames the issue of combining predictions as a decision theory problem - a decisionmaker rationally synthesizing some set of information to inform her choice of action. The theoretical underpinnings are provided by West (1992) and West and Crosse (1992) who shows how a decisionmaker 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 function form, or 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 have used a dynamic linear model as a synthesis function (Prado and West, 2010, Sect. 4.5). Instead, BPS has been 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. (2022) 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 numbers of parameters with small datasets. From a frequentist perspective the approach has been 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 likelihood for each model which is computationally expensive. Researchers have addressed this issues 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 will tend to place weight on a smaller subset of experts (sparsity) while factor models will look for co-movement and the weights will be 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 which have examined 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.

Specifically, I use the triple gamma prior (Cadonna et al., 2019) 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 models the weights as correlated with a factor structure which is in contrast to this paper where the forecasts are modelled as correlated and have a factor structure.

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 density combinations in economics[1]. Several authors have shown that combining densities can make predictions more robust and improve their accuracy (Jore et al., 2010; Del Negro et al., 2016), while others have worked on specifying optimal combination strategies from both frequentist, (Conflitti et al., 2015) and Bayesian perspectives (Geweke and Amisano, 2011). More recent academic work focuses on modelling the dependence and correlation across forecasts, and time variation in weights[2]. Furthermore, Knotek and Zaman (2020) 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).

Summarizing the previous discussion, the main contribution of this paper is to compare the performance of global-local shrinkage priors and factor models as synthesis functions within the BPS framework. This is interesting because comparisons of BPS synthesis functions have not been studied and the functions considered in this paper address the problem of combining large numbers of forecasts. More broadly, I contribute to the literature on density combinations by documenting empirical regularities in two common applications macroeconomic forecasting. In particular, I consider nowcasting Canadian GDP with model based forecasts and combining survey forecasts of Euro-Area GDP from the European Survey of Professional Forecasters.

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 is on 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 some cases, time-varying parameter specifications can reduce to a time-varying mean model

[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)

which overfits the model resulting in poor out-of-sample performance.

The remainder of the paper proceeds as follows: section 2 describes Bayesian predictive synthesis along with an outline of the MCMC approach. This is followed by a description of the forecast combination techniques - the synthesis functions. In addition, I will provide a brief overview of global-local shrinkage priors and Bayesian factor models. Section 3 the details the prediction exercises before section 4 discusses the results. Section 5 concludes.

2 Econometric Framework

This section describes the econometric framework used in this paper. Since BPS is a framework for combining density predictions, I will start with describing BPS and move into descriptions of the synthesis functions. Appendix A has details on the global-local shrinkage priors, implementation of the factor model combination, sampling of x_t, and the an overview of the MCMC algorithm to estimate the density combinations.

2.1 Bayesian Predictive Synthesis

The basic idea and implementation of BPS is straightforward[3]. Given a set of forecast densities, $h_j(x) \in \mathcal{H}$, we want to find a distribution of the target variable (y) conditional on these densities: $p(y|\mathcal{H})$. In other words, given a set of forecasts, for say GDP, we want to find the distribution of GDP conditional on those forecasts. Applying Bayes Theorem implies a straightforward MCMC routine with two steps. First, estimate a 'synthesis' function to combine the forecasts - $\alpha(y|x)$. This synthesis function is estimated on (conditional) draws from the forecast distributions (x). The second step is to draw iterates from the forecast distributions *conditional* on the target variable (GDP). Key to understanding BPS is the interpretation of x_t. McAlinn and West (2019) describe x_t as a factor, but as Aastveit et al. (2022) point out x can be thought of as a generated regressor. This means BPS can be thought of as a multivariate regression model with generated regressors as predictors.

More formally, the decisionmaker $\mathcal{D}$ is presented with $h_j(x) \in \mathcal{H}$, where $h_j(x)$ is the set of

[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

density functions that are elements of the information set $\mathcal{H}$. The agent opinion analysis theory (West (1992) and West and Crosse (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 $x_t = x_{t,1:J} = (x_{t,1}, \ldots, x_{t,J})'$ is a J-dimensional vector and $\alpha(y_t|x_t)$ is a conditional probability distribution function for y_t given x_t. Where Φ_t are the synthesis function parameters. This equation shows how to relate a set of agent forecast distributions to the decisionmakers combined forecast or, in more simple terms, how to combine forecast distributions in a Bayesian fashion. With equation 1 in hand we 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):

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

Where y_t is the target variable, x_t are draws from the forecast distributions (including a vector of ones for intercept), β_t are combination weights that vary over time following a random walk with variance θ, and ϵ_t an error term exhibiting stochastic volatility with volatility σ_t^2.

The first step in BPS would be 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 has to specify the function and estimate 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

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

If the individual expert densities are normal, this yields a multivariate normal for x_t. However, the applications in this paper do not have analytical distributions. For example, the European SPF elicits histograms from survey respondents. Since there is no analytic representation of the densities, we cannot derive a conditional density to sample from. Instead, I use a block Metropolis-Hastings step to sample the x_t using the aforementioned multivariate normal as a proposal distribution. Details are provided in the appendix.

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. (2019)) 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.

It is possible to impose shrinkage on time-varying parameter models such as equation 2 by

rewriting it 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. (2019) 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 we 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.

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

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

Where f_t 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 $\sigma_{\nu J}^2$. In order to derive combination weights we need to identify the factors. This is done by using the following restriction $f_t' f_t = I_J$ and 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.

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 we need to evaluate $p(y_t|x_t\beta_t, \epsilon_t)$ during the MH step but equation 5 is specified in terms of f_t. However, the model can be reparametrized in terms of x_t and $x_t|y_t, \Phi_t$ and sampled using the standard technique.

3 Forecasting Environment

This section describes the two environments in which the synthesis functions will be assessed. The first exercise is nowcasting Canadian real GDP in pseudo real-time using a large set of models, while the second exercise is forecasting Euro Area real GDP using the Survey of Professional Forecasters. These two environments are very different: not only are they in different regions, but the types of forecasts provided are different. The nowcasting exercise uses model based predictions from four different model classes. While the forecasting exercise features mostly judgemental forecasts (Bank (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 a comprehensive assessment of the various synthesis functions.

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 reader can consult these papers and references within for detailed results and descriptions of the models. In short, Chernis and Webley (2022) produces density nowcasts for Canadian GDP using four commonly used model classes (see figure 1) totalling 98 models. Pseudo-real-time forecasts are produced from 2000 to 2021 and real-time predictions from 2013 to 2019. In this paper we use the pseudo real-time forecasts with a

5 year expanding estimation window, the training sample goes from 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, and 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 2 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 by 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 Garca (2003). The quarterly survey began in 1999 and is the longest running Euro area survey of macroeconomic expectations. The survey elicits probability and point forecast on inflation and GDP growth at various horizons (we 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).

There are several attributes of the survey that 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% and increasing by 50 basis point intervals to 4% 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 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 this issue.

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 is a large amount of missing forecasts, I drop forecasters with less than 5 forecasts in a 5 year period. This 5 year period is also the rolling estimation window we use for the model. After dropping forecasters for each five year period we are left with an unbalanced panel that has about 35 respondents each quarter. One consequence of the changing panel composition is that examining the online weights is not meaningful because there will be 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[4]. 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

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

'wide' dataset since there is much less missing data being imputed and a longer timeseries to train the algorithm.

Once the data set has been assembled the first estimation window is 1999Q3 to 2004Q2, and the evaluation window is 2005Q2 to 2020Q4. The forecast combination is estimated with a 5-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

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

I begin assessing the different synthesis function by examining the average CRPS for all variants of the shrinkage and the factor-based combination approaches in both the nowcasting and survey forecast combination applications. Figure 3 shows the results for the nowcasting exercise and figure 4 for the survey forecast application. The heatmaps are shaded such that lighter shading corresponds to lower CRPS and darker shading to higher CRPS. There are very similar results across both exercises. Comparing the shrinkage approaches to the factor model based combination, one can see that, in general, shrinkage priors have lower CRPS. There are some exceptions; notably, the LASSO performs poorly, shrinkage performs worse at the forecast horizons (24 to 22 before the National Accounts), and in the 'tall' dataset, shrinkage and factor models with constant parameters are competitive. However, these latter two exceptions can be explained by the COVID-19 pandemic and idiosyncratic factors.

Examining the cumulative CRPS difference will help us to understand differences in relative forecasting performance by showing how relative forecast accuracy changes over time. This is particularly useful for highlighting episodes that may have undue influence on average forecast accuracy. For brevity, I will 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 2 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 some commonalities that will become apparent as we analyse the results. In both applications, the triple gamma prior performs slightly worse at the beginning of the sample, but 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. There is an exception to this in figure 6. During the Euro Area crisis, the factor model improves over the triple gamma approach. It turns out that this is because one of the forecasters drops out from the sample for three quarters, and it so happens they had a large proportion of the weight. The factor 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. The next important event is the COVID-19 pandemic, which has a significant impact on the relative model performance and is behind the sometimes competitive performance of the factor model combination. 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 20Q1 and 20Q2 more accurately, but it misses the sharp and immediate rebound in 20Q3, and the blue line approaches the y-axis signifying that on average the two methods perform similarly. There is a similar result for the survey forecast application, the factor model is relatively more accurate than the triple gamma prior. This is not so much due to the factor model providing a significantly better forecast, since it missed by a large margin, rather the variance of the forecast distribution is slightly higher such that those combinations were punished less for inaccurate predictions.

4.2 Predictive Accuracy: Time-varying and Constant Parameter Combinations

Another key finding is that constant parameter combinations generally have a lower CRPS than their time-varying counterparts. Comparing the columns in figures 3 and 4 for time-varying parameter combinations with their constant parameter counterparts shows there can be significant gains for choosing a more parsimonious specification. In the nowcasting application there are gains 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 on 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 7 shows the in-sample time-varying intercept for the triple-gamma prior and the one-factor combination approach overlaid with the 4-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, if we inspect the weights for each of the time-varying combination methods we can see that the weight put on individual experts is quite small (lower panel in figure 8). In contrast, the upper panel of figure 8 shows that the sum of weights from constant parameter specifications is much closer to 1. Taken together, we can see that the poor performance of the time-varying parameter models is that 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. (2022)) which have found 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 has large and persistent movements in the price that makes a time-varying intercept useful. In contrast, Euro Area real GDP has much smaller movements in its growth rates over the 20 year period in question.

4.3 Examining the Combination Weights

It is instructive to examine the weights in figure 8 to gain some intuition on the implications of synthesis function choice. Let us start with the weights from the triple gamma prior in the top left panel.

First, we can see that the combination method puts significant weight on a single expert, a handful of other forecasts, and close to zero weight on the rest[5]. This prior implies the decision

[5]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

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 which 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 decisionmaker hedges against the high weight on a given expert by 'short-selling' a similar correlated forecast.

Second, if we examine the right panel we can see that BFM-1 has weights that are spread more evenly over experts (but not equally) meaning the combination is closer to consensus weights[6]. 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 follows a consensus based approach to processing forecasts, and the approach is quite different from shrinkage priors where the decisionmaker focuses on a small subset of experts. The results above suggest that sparse weights are preferable to consensus weights - a decisionmaker should not follow the herd, but instead focus on a smaller set of experts.

In the appendix we assess the absolute accuracy of the combination methods by testing if they are correctly calibrated over the forecast evaluation period using the test of Knüppel (2015). The results suggests that forecasts from both synthesis functions are well calibrated and provide a good approximation of the data distribution.

5 Conclusion

In this paper, I investigated different approaches to 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 forecast, and combining large numbers of forecasts requires specialized modelling techniques. I used two common approaches in economics to deal with large datasets: global-local shrinkage priors and factor modelling. In particular, I used the newly developed triple gamma

worse forecasting performance suggesting the non-zeros weights are not numerical artifacts.

[6] 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

prior, and the priors it encompasses, along with a novel approach to forecast combinations: a Bayesian Factor Model. The approaches are tested 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 which allows 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 is an important finding as recently developed combination schemes tend to utilize time-varying parameter specifications. Second, and more importantly, I find that in general shrinkage approaches 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. This is interesting as the two synthesis functions imply very different weighting structures. The sparse weighting scheme of shrinkage priors implies that decisionmakers should give considerable weight to a smaller set of experts. In contrast, the factor model based scheme extracts the co-movement between the predictions. Therefore, my results suggests that focusing on a smaller set of accurate experts is preferable to following a crowd.

References

Aastveit, K. A., Cross, J. L., and van Dijk, H. K. (2022). Quantifying Time-Varying Forecast Uncertainty and Risk for the Real Price of Oil. *Journal of Business & Economic Statistics*, 0(0):1–15. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/07350015.2022.2039159.

Aastveit, K. A., Gerdrup, K., and Jore, A. S. (2011). Short-term forecasting of GDP and inflation in real-time: Norges Bank's system for averaging models. page 56.

Aastveit, K. A., Mitchell, J., Ravazzolo, F., and Dijk, H. v. (2018). The Evolution of Forecast Density Combinations in Economics. Technical Report 18-069/III, Tinbergen Institute.

Aastveit, K. A., Ravazzolo, F., and Dijk, H. K. V. (2016). Combined Density Nowcasting in an Uncertain Economic Environment. *Journal of Business & Economic Statistics*, 0(0):1–15.

Bache, I. W., Mitchell, J., Ravazzolo, F., and Vahey, S. P. (2009). Macro modelling with many models. Technical Report 2009/15, Norges Bank.

Bank, E. C. (2019). Results of the third special questionnaire for participants in the ECB Survey of Professional Forecasters.

Belmonte, M. A. G., Koop, G., and Korobilis, D. (2014). Hierarchical Shrinkage in TimeVarying Parameter Models. *Journal of Forecasting*, 33(1):80–94. Publisher: John Wiley & Sons, Ltd.

Billio, M., Casarin, R., Ravazzolo, F., and van Dijk, H. K. (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, H. C., Gerdrup, K., Jore, A. S., Smith, C., and Thorsrud, L. A. (2012). Does Forecast Combination Improve Norges Bank Inflation Forecasts? *Oxford Bulletin of Economics and Statistics*, 74(2):163–179.

Brown, P. J. and Griffin, J. E. (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. (2019). Triple the gamma – A unifying shrinkage prior for variance and variable selection in sparse state space and TVP models. *arXiv:1912.03100 [econ, stat]*. arXiv: 1912.03100.

Carvalho, C. M., Polson, N. G., and Scott, J. G. (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, H. K. (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, J. C. and Jeliazkov, I. (2009). Efficient simulation and integrated likelihood estimation in state space models. *International Journal of Mathematical Modelling and Numerical Optimisation*, 1(1/2):101.

Chernis, T. and Sekkel, R. (2017). A dynamic factor model for nowcasting Canadian GDP growth. *Empirical Economics*, 53(1):217–234.

Chernis, T. and Sekkel, R. (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, J., 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. Publisher: Elsevier.

Del Negro, M., Hasegawa, R. B., and Schorfheide, F. (2016). Dynamic prediction pools: An investigation of financial frictions and forecasting performance. *Journal of Econometrics*, 192(2):391–405.

Diebold, F. X., Gunther, T. A., and Tay, A. S. (1998). Evaluating Density Forecasts with Applications to Financial Risk Management. *International Economic Review*, 39(4):863–883.

Diebold, F. X., 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.

Fava, B. and Lopes, H. F. (2021). The illusion of the illusion of sparsity: An exercise in prior sensitivity. *Brazilian Journal of Probability and Statistics*, 35(4):699–720. Publisher: Brazilian Statistical Association.

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.

Garca, J. A. (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, G. E. (2021). Economic Predictions With Big Data: The Illusion of Sparsity. *Econometrica*, 89(5):2409–2437. Publisher: Econometric Society.

Gneiting, T. and Raftery, A. E. (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, S. G. and Mitchell, J. (2007). Combining density forecasts. *International Journal of Forecasting*, 23(1):1–13.

Hartkopf, J. (2022). gigrnd.

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

Jore, A. S., Mitchell, J., and Vahey, S. P. (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, H. F. (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, E. S. and Zaman, S. (2020). Real-Time Density Nowcasts of US Inflation: A Model-Combination Approach. Technical Report 20-31, Federal Reserve Bank of Cleveland. Publication Title: Working Papers.

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, H. F. (2014). Modern Bayesian Factor Analysis. In Jeliazkov, I. and Yang, X.-S., editors, *Bayesian Inference in the Social Sciences*, pages 115–153. John Wiley & Sons, Inc., Hoboken, NJ, USA.

Lopes, H. F. 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. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/rssc.12500.

McAlinn, K., Aastveit, K. A., 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, W. J., Miller, S., and Pelletier, D. (2011). Simulation smoothing for statespace 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, A. E. (2016). Dynamic model averaging in large model spaces using dynamic Occams window. *European Economic Review*, 81:2–14.

Prado, R. and West, M. (2010). *Time Series: Modeling, Computation, and Inference*. Chapman and Hall/CRC, Boca Raton, 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.

Rue, H. and Held, L. (2005). Gaussian Markov Random Fields: Theory and Applications.

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

Wallis, K. F. (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, X.-L. (2011). To Center or Not to Center: That Is Not the QuestionAn AncillaritySufficiency 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

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 1: Model List

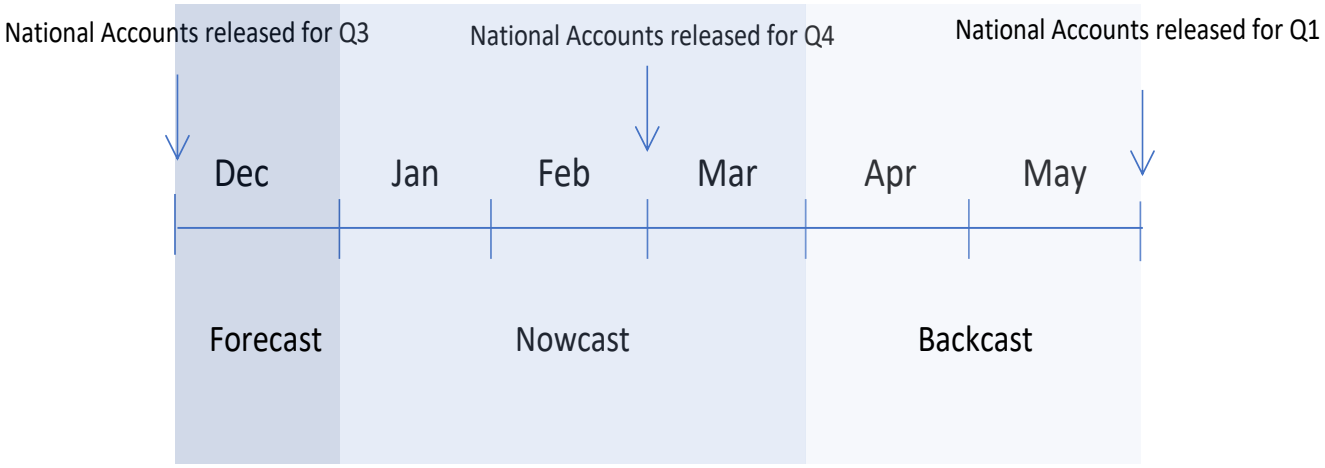

Figure 2: Overview of forecast cycle

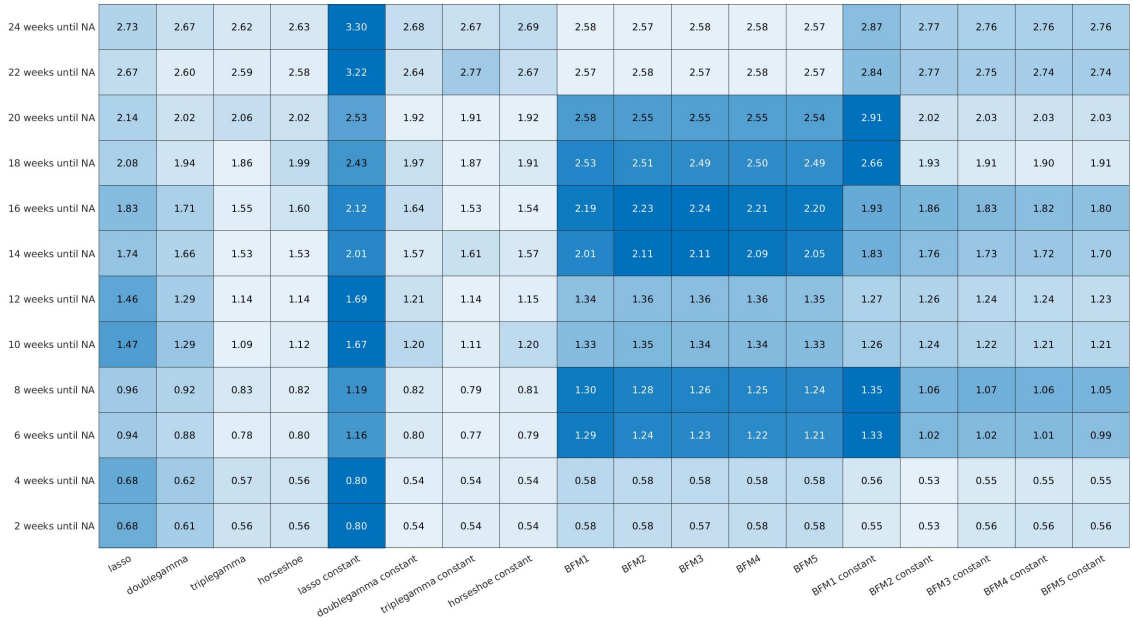

Figure 3: CRPS in Nowcasting Exercise

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. Lighter shading corresponds to lower CRPS in that row and darker shading to higher CRPS

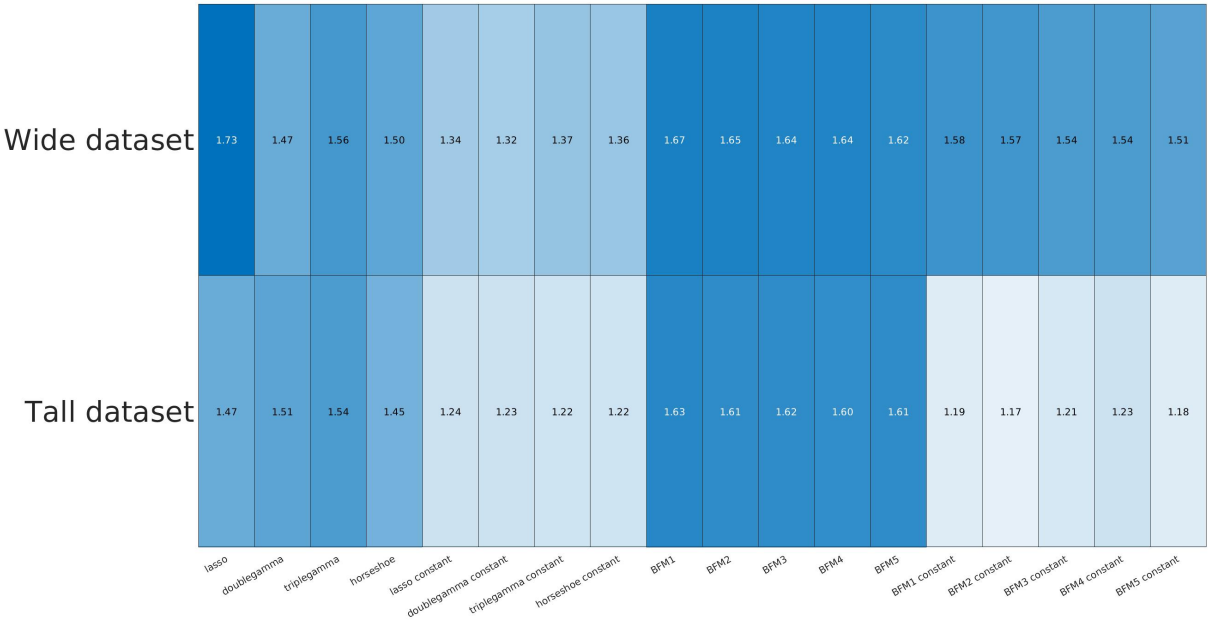

Figure 4: Continuous Rank Probability Score for SPF forecasts

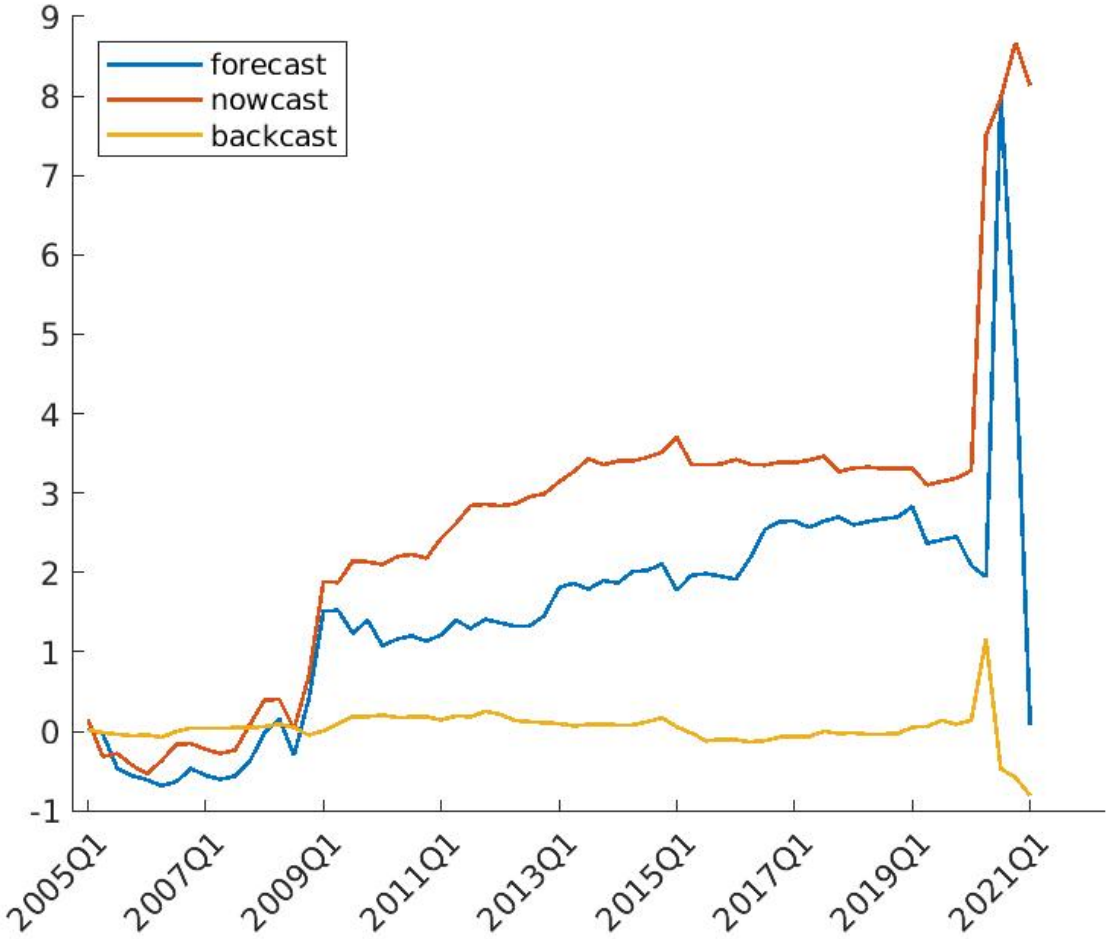

Figure 5: Nowcasting Cumulative CRPS difference: Factor Model 2 (constant) - Triple Gamma (constant)

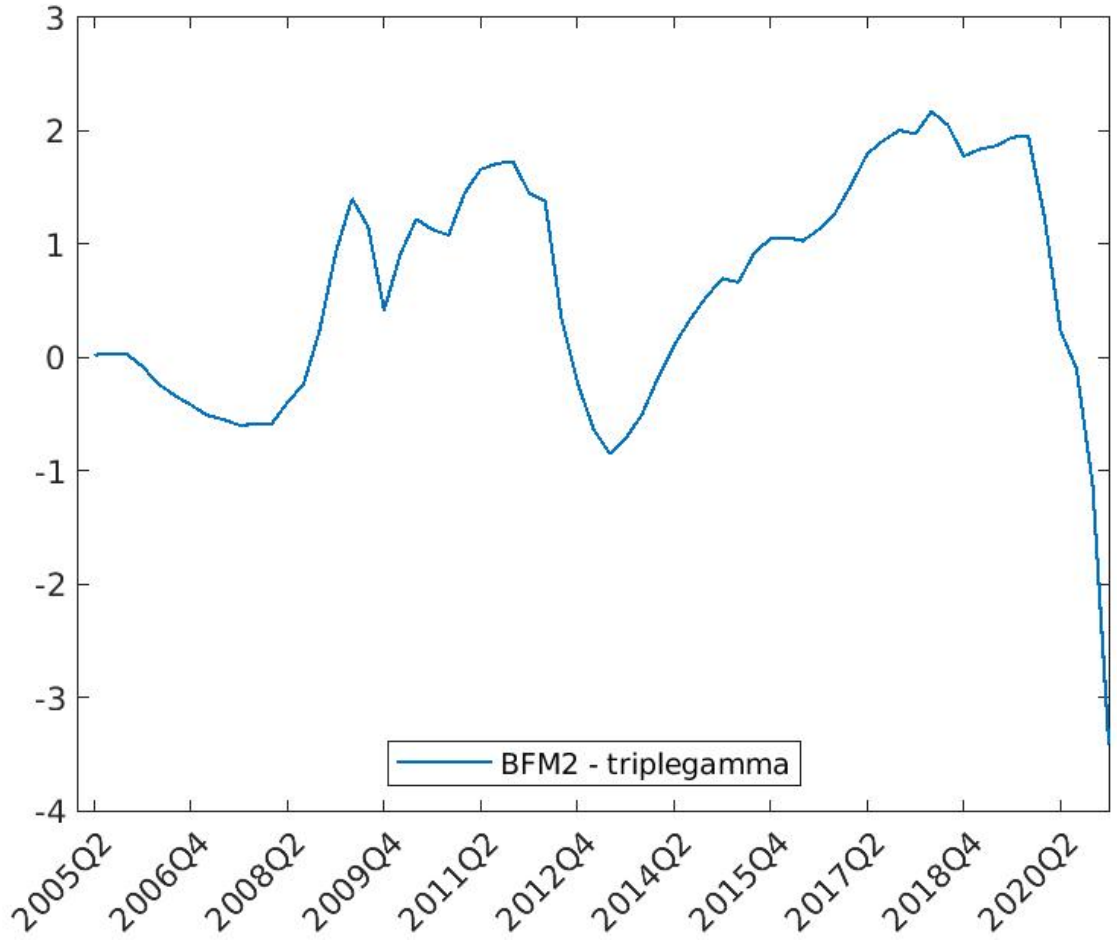

Figure 6: SPF Cumulative CRPS difference: Factor Model 2 (constant) - Triple Gamma (constant)

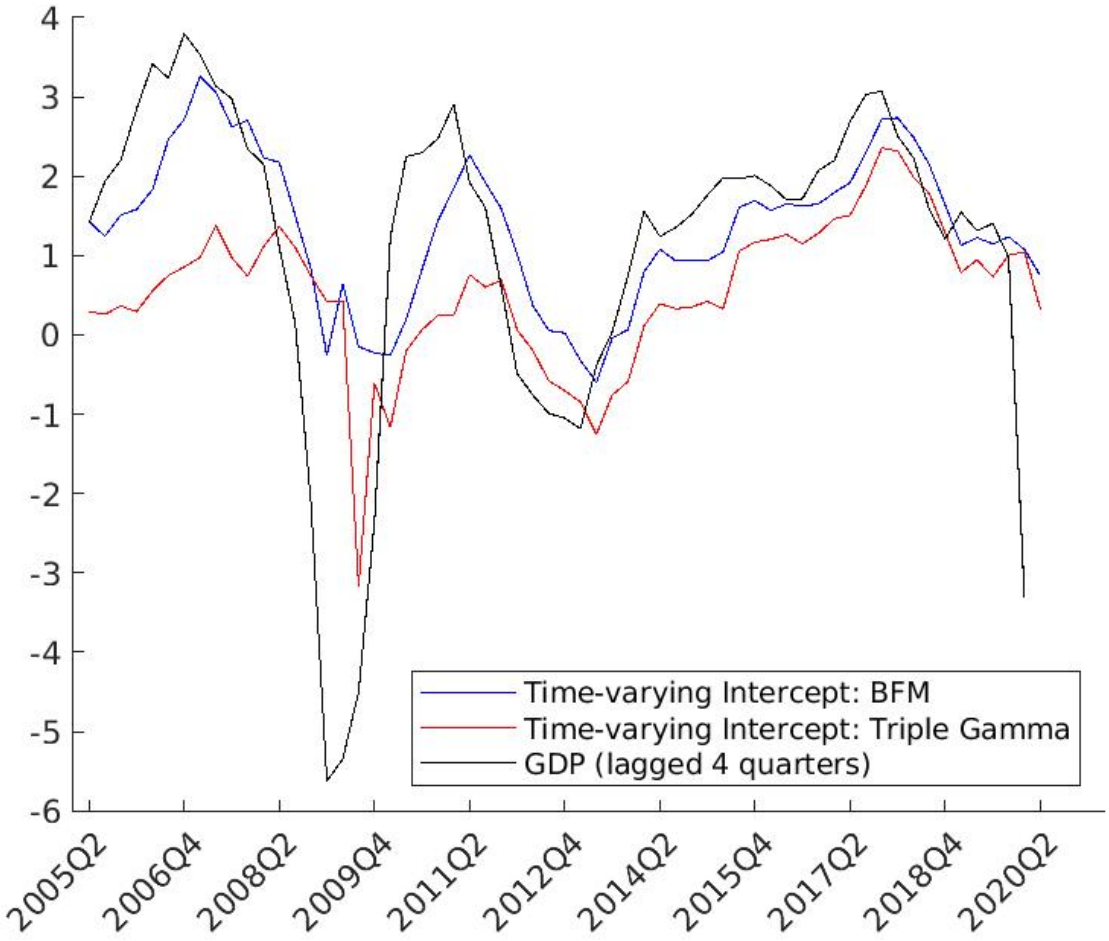

Figure 7: Time-varying predictive of mean of BPS intercepts

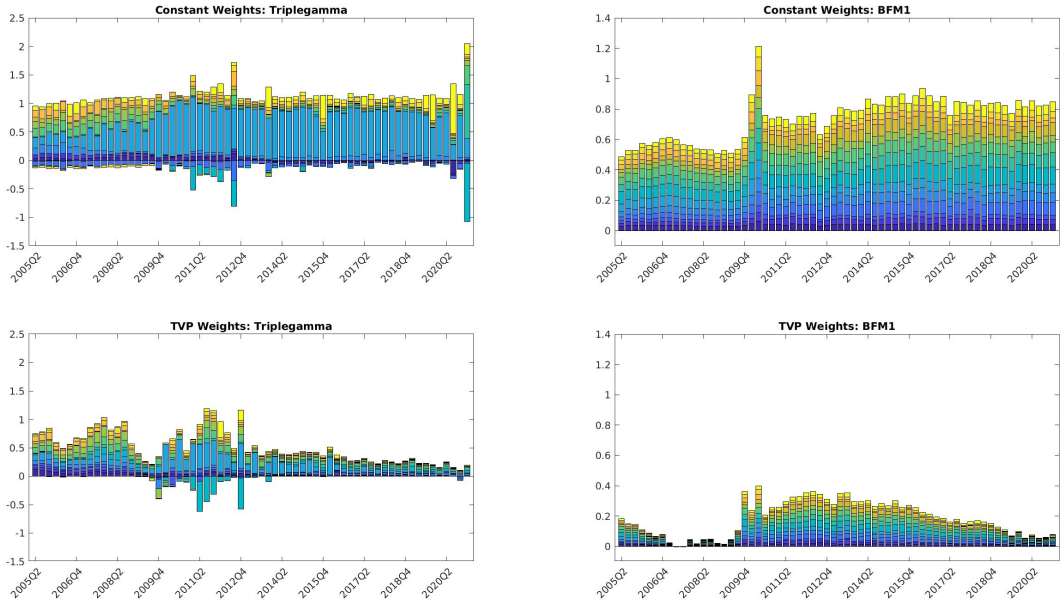

Figure 8: Sequentially estimated mean combination weights

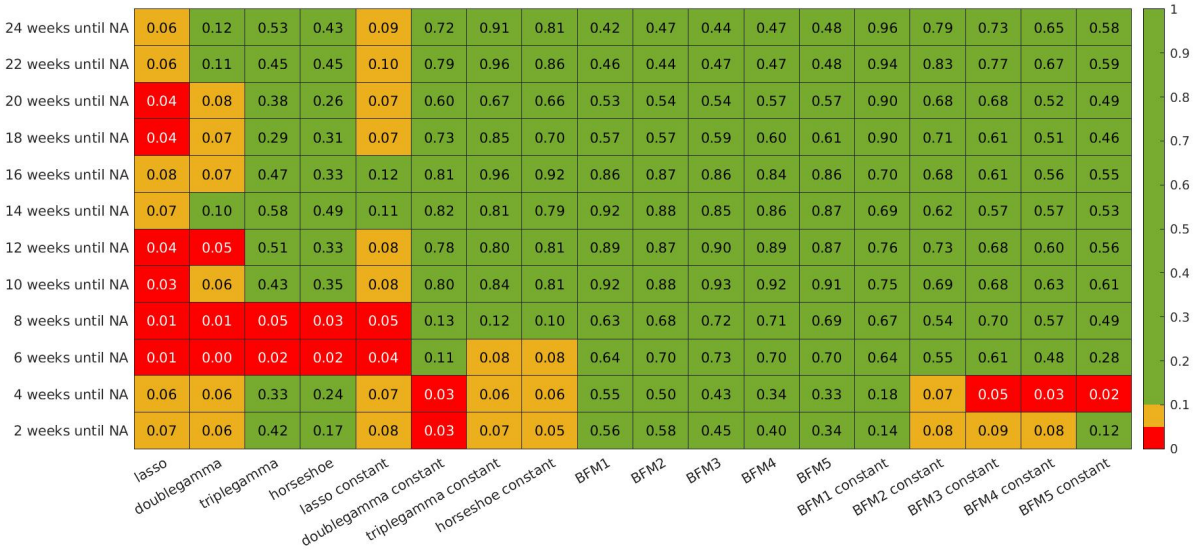

Figure 9: Knüpple test for probabilistic calibration: Nowcasting Application

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 per cent level and yellow at 10 per cent level.

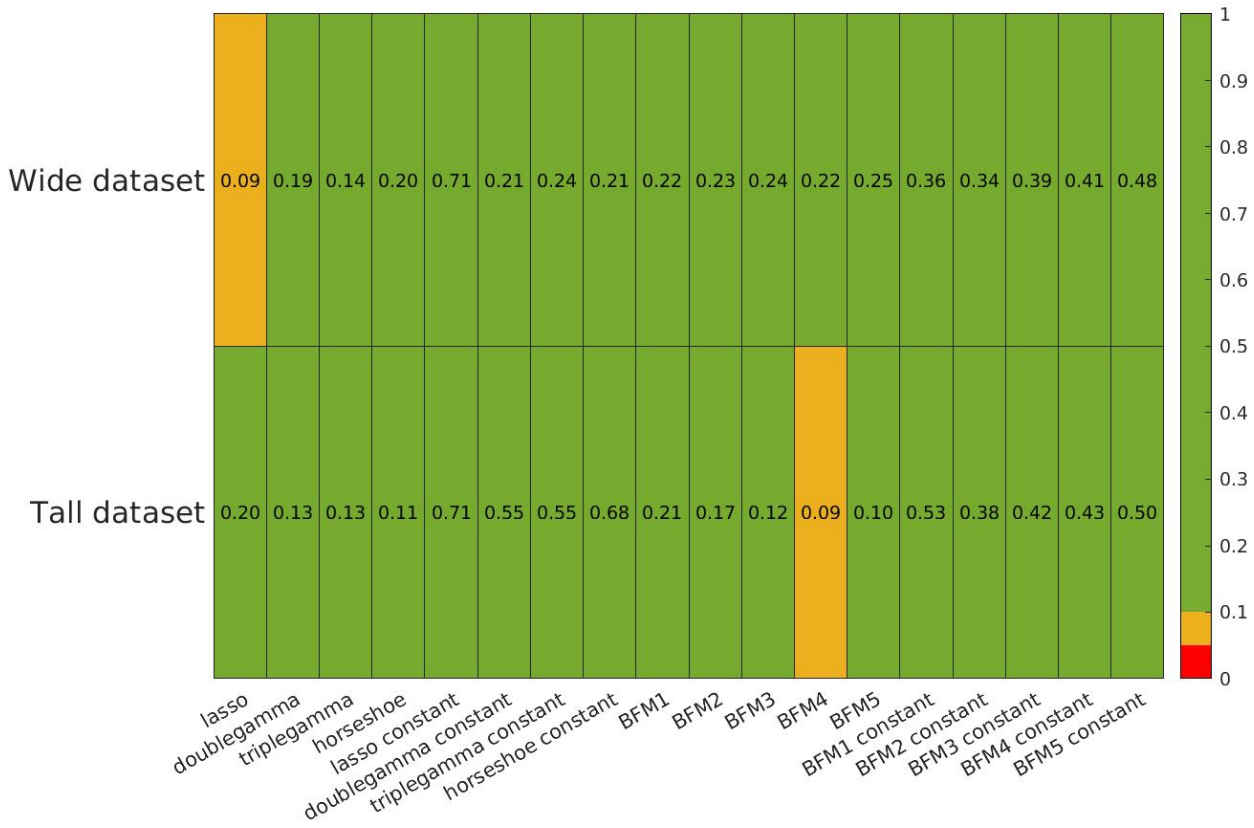

Figure 10: Knüpple test for probabilistic calibration: Survey Forecast Application

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 per cent level and yellow at 10 per cent level.

A Technical Appendix

A.1 MCMC algorithm

This section describes the Markov Chain Monte Carlo algorithm used to estimate the forecast combinations. It largely follows McAlinn and West (2019) for the BPS steps, Cadonna et al. (2019) 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 will discuss 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) provides 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. (2019) and Bitto and Frühwirth-Schnatter (2019). First, I will 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 $x_t = (x_{t1}, x_{t2}, \dots, x_{td})$ is a d-dimensional row vector containing the regressors at time t, with x_{t1} corresponding to the intercept.

For simplicity, we assume here that $Q = \text{Diag}(\theta_1, \dots, \theta_d)$ is a diagonal matrix, implying that the state innovations are conditionally independent. Moreover, we 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,

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

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 model. 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., 2019), 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 we place a 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. (2019) 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. (2019) 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, we 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 $\alpha_{a^{\xi}}$ and $\alpha_{a^{\tau}}$ act as degrees of freedom and allow the prior to be bounded away from zero.

In the constant parameter case we 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 and for more details please refer to Cadonna et al. (2019) and Bitto and Frühwirth-Schnatter (2019) for further details.

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. 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 (c) - (f) from Algorithm 1 in Cadonna et al. (2019). In the NG case steps (d) and (e) from Algorithm 1 in Bitto and Frühwirth-Schnatter (2019) are used.
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. (2019). 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. We 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. The reader is referred to those references for detailed discussion on the methods. Here we provide a brief overview of the model and estimation technique.

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

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

Where, f_t is a $K \times 1$ vector of factors and define $F = (f_1, \ldots, f_t)'$ with F_i as the $T \times i$ matrix containing the first i columns of F, β_t is $K \times 1$ 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 we 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 we 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 variances are independent and modelled 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 β_t are $\beta_0 \sim \mathcal{N}(0, P0)$ where $P0 \sim \mathcal{IG}(\nu_P, (\nu_P - 1) * c_P)$.

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 $C_i^{-1} = C_0^{-1} I_i + \sigma_{\nu i}^2 F_i' F_i$;
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 $\beta = (\beta_0, \ldots, \beta_T)$ from a multivariate normal distribution. Otherwise, sample jointly $\beta_1, \ldots, \beta_d$, in from a multivariate normal distribution using the precision sampler of Chan and Jeliazkov (2009).
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_η^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 $c_t = y_t - \beta_t 0 - h_t'\beta_{t,1:J}$, $g_t = \sigma_t^2 + \beta_{t,1:J}' H_t \beta_{t,1:J}$ and $b_t = H_t \beta_{t,1:J}/g_t$. Unfortunately, the applications in this paper do not have analytical forms instead we have histograms representing the agent densities. With no analytical form we use a Block Metropolis-Hastings step with 24 as a proposal distribution. Since the number of agent densities can be large we break the MH step into blocks of 5 experts which are sampled at a time.

There are a few details for Bayesian Factor Model combinations which warrant explanation. First, is that the model has to re-parameterized in terms of the x_t so that we can use the proposal distribution from 24 in the MH step. The model is straightforward to re-parameterize with the following steps:

$$x_t = \Lambda f_t + \nu_t \tag{25}$$

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

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

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

Now that the model has been re-parameterized we can use the equation 24 in the MH step by substituting in $\beta_t = \beta_t^*$, and error variance $\epsilon_t^* \sim \mathcal{N}(0, \beta_t'\Omega R\Omega'\beta_t + \sigma_t^2)$.

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 mean and variance used to standardized draws from the agent densities is calculated 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

In this section we assess 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 per cent probability should be observed in the data roughly 20 per cent 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). We assess calibration with test based off 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 9 and 10 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 has 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 which calibration is rejected at the 10 per cent 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 per cent level.