

ISSN: (Print) (Online) Journal homepage: <https://www.tandfonline.com/loi/ubes20>

# Using Survey Information for Improving the Density Nowcasting of U.S. GDP

Cem Çakmaklı & Hamza Demircan

To cite this article: Cem Çakmaklı & Hamza Demircan (2022): Using Survey Information for Improving the Density Nowcasting of U.S. GDP, Journal of Business & Economic Statistics, DOI: [10.1080/07350015.2022.2058000](https://doi.org/10.1080/07350015.2022.2058000)

To link to this article: <https://doi.org/10.1080/07350015.2022.2058000>



View supplementary material [↗](#)



Published online: 27 Apr 2022.



Submit your article to this journal [↗](#)



Article views: 203



View related articles [↗](#)



View Crossmark data [↗](#)



# Using Survey Information for Improving the Density Nowcasting of U.S. GDP

Cem Çakmaklı<sup>a</sup> and Hamza Demircan<sup>b</sup>

<sup>a</sup>Department of Economics, Koç University, Istanbul, Turkey; <sup>b</sup>Central Bank of the Republic of Turkey, Ankara, Turkey

## ABSTRACT

We provide a methodology that efficiently combines the statistical models of nowcasting with the survey information for improving the (density) nowcasting of U.S. real GDP. Specifically, we use the conventional dynamic factor model together with stochastic volatility components as the baseline statistical model. We augment the model with information from the survey expectations by aligning the first and second moments of the predictive distribution implied by this baseline model with those extracted from the survey information at various horizons. Results indicate that survey information bears valuable information over the baseline model for nowcasting GDP. While the mean survey predictions deliver valuable information during extreme events such as the Covid-19 pandemic, the variation in the survey participants' predictions, often used as a measure of "ambiguity," conveys crucial information beyond the mean of those predictions for capturing the tail behavior of the GDP distribution.

## ARTICLE HISTORY

Received October 2020  
Accepted March 2022

## KEYWORDS

Bayesian inference;  
Disagreement; Dynamic  
factor model; Predictive  
density evaluation;  
Stochastic volatility; Survey  
of professional forecasters

## 1. Introduction

Monitoring economic conditions in a timely and accurate manner is crucial for economic agents. Gross domestic product (GDP), however, as the key indicator of the economic conditions, is not available instantly as it is released with a substantial delay. Two sources of GDP projections are typically available for the decision maker. First, econometric models of nowcasting have been proved to be very useful in providing accurate predictions of the GDP density using a large set of economic and financial indicators that are timely available. Second, surveys, reporting predictions of forecasters, serve as an important guide reflecting prompt reactions of these forecasters to changing economic conditions.

Econometric specifications constructed for nowcasting often relies on dynamic factor models, among others,<sup>1</sup> to extract the common movement in the economy, which in turn, is used for the prediction of the current GDP. Dynamic factor models have the advantage of processing datasets composed of a large number of macroeconomic and financial variables sampled at mixed frequencies in a statistically optimal manner to extract the common behavior in these series or put differently, factors, see, for example, Giannone, Reichlin, and Simonelli (2009), Banbura et al. (2013), and Bok et al. (2018). Coupled with the time-varying volatility, factor models constitute the workhorse for predicting the current economic activity and its distribution,

see, for example, Marcellino, Porqueddu, and Venditti (2016) for a recent study.

Surveys, on the other hand, reflecting the expectations of the key economic agents, bear specific information that either based on the specific model or the judgment of the survey participants about future economic conditions. Essentially, survey participants form their "judgmental" expectations based on both public and private information they might have. Ang, Bekaert, and Wei (2007), Campbell (2007), D'agostino, McQuinn, and Whelan (2012), and Faust and Wright (2013), among others, document that surveys indeed convey useful information beyond that is provided by the statistical models. Following the great recession of 2008–2009, a recent strand of research focuses on the ambiguity reflected in the predictions of survey participants as a proxy for uncertainty especially during turmoil periods, see, for example, Giordani and Söderlind (2003), Lahiri and Sheng (2010), Bloom (2014), and Grischenko, Mouabbi, and Renne (2019). Therefore, we distinguish two important features of survey-based information. First, survey participants provide useful information that might diverge from the statistical models particularly at times of high uncertainty. Second, disagreement among survey participants serves as an important proxy of ambiguity those agents confront.

Our point of departure in this article is the construction of an efficient combination of the baseline statistical model with the survey information for exploiting the distinct features of these two approaches. First, we depart from conventional models of nowcasting based on the dynamic factor model and we augment it with stochastic volatility components for the GDP and for the factor structure. By doing this, we construct a competitive baseline statistical model for nowcasting the density of the GDP using the time variation in the mean and the volatility. Second,

<sup>1</sup>Other popular methods used for nowcasting the GDP involve mixed frequency VARs, see, for example, Giannone, Reichlin, and Simonelli (2009) and Foroni, Ghysels, and Marcellino (2013), among others, or MIDAS and bridge equations, see, for example, Aastveit et al. (2014) and Schumacher (2016), among others, which make use of datasets at mixed frequencies.

our methodology of integrating the survey information into the statistical model relies on the fact that survey information represents the predictions formed by the survey participants. Hence, our combination strategy aligns the predictions implied by the statistical model with the predictions of the survey participants. We do this by matching both the mean and the variance of the predictive distributions obtained from the statistical model together with the mean and the variance of the survey expectations. This corresponds to augmenting the baseline model with new measurement equations representing these alignments.

We evaluate a dynamic factor model (using a dataset comprised of 17 variables) with and without survey information and stochastic volatility components based on a horse-race in terms of various measures of density evaluation. In particular, we employ likelihood and squared forecast error based metrics to evaluate the full-sample and out-of-sample performance of competing models with a special focus on the tail behavior. Considering baseline statistical models, we observe that adding stochastic volatility over the conventional baseline model with constant volatility does not improve the model fit but it improves nowcasting and forecasting performances. On the other hand, when we combine statistical models with the survey information, obtained from the Survey of Professional Forecasters (SPF), the model performance improves substantially. While, using only the mean of the survey expectations improves upon the baseline model, incorporating the variance of the survey expectations provides a further improvement. Interestingly, survey expectations at all horizons including the survey nowcasts of the current quarter and forecasts up to 4-quarter ahead bear additional information beyond that is provided by the baseline statistical model. These findings are robust for both nowcasts and short- and medium-horizon forecasts up to two quarters ahead.

We further demonstrate the efficacy of the proposed framework by focusing on the month-by-month performance of the selected models during the six months of 2020 starting from April, that is, during the exceptional recent periods when the Covid-19 pandemic hit the U.S. economy and the following rapid bounce-back periods. In this case, while the baseline statistical model predicts relatively milder downturn and reversal than the extreme realizations in the second and third quarters, the survey forecasters' mean predictions turn out to be quite accurate. Combining these two sources, the proposed model outperforms the baseline statistical model clearly both in terms of point forecast and density forecast. These demonstrate the valuable information provided by the survey participants during turbulent times.

Our approach is closely related to the approach followed by Grischenko, Mouabbi, and Renne (2019) who combines the predictions from an unobserved components model together with the information in survey-based expectations. They do so for measuring the inflation uncertainty and whether inflation expectations are anchored in the United States and Euro Area by taking this uncertainty into account, see also Kozicki and Tinsley (2006, 2012) and Altug and Çakmaklı (2016) who use only the first moment of survey expectations. Aruoba (2020) derives the term structure of inflation expectations by matching the first moment of the survey-based inflations with the predictions of a dynamic factor model, specifically the Nelson–

Siegel model. A related line of research uses survey information to combine it with the forecasts based on Bayesian Vector Autoregressions (BVAR) involving a relatively limited number of variables, typically using entropic tilting or forecast combination methods. Krüger, Clark, and Ravazzolo (2017), for example, employ entropic tilting using both the first and the second moments for matching the medium-term forecasts from a BVAR with the nowcasts from surveys, see also Doh and Smith (2018) and Tallman and Zaman (2019) for similar approaches. Altavilla, Giacomini, and Ragusa (2017) use entropic tilting for combining survey-based expectations with the predictions of the Nelson–Siegel model for predicting the yield curve. Our approach, on the other hand, focuses on nowcasting and short-term forecasting of the density of the U.S. real GDP by using a dynamic factor model using a dataset with 17 variables with mixed frequency together with the survey-based information on the first and second moments of the real GDP growth distribution. Alternatively, Billio et al. (2013) and McAlinn and West (2019) provide dynamic forecast density combination methods. These methods combine (the moment(s) of) the predictive distributions obtained from the plain statistical model together with the (moment(s)) of the predictive distribution from the additional source of information ex-post. On the contrary, our approach yields a predictive distribution by incorporating the additional source of information into the statistical model structure ex-ante. We do this by extending the model structure by including additional measurement equations using the forward-looking behavior of survey-based expectations. This unified approach also leads to significant computational efficiency. We provide a comprehensive account of comparing our methodology with these methods. Results show that our method leads to accurate predictive densities, and it reacts to changing conditions swiftly compared to these alternative methods.

The remainder of this article is as follows. Section 2 presents model specifications. Section 3 discusses the data, details on estimation methodology and model evaluation.<sup>2</sup> Section 4 evaluates the empirical results. Section 5 discusses the performance of competing models during the Covid-19 pandemic recession. Section 6 provides a detailed comparison between alternative model strategies. Finally, Section 7 concludes.

## 2. Model Details

In this section, we present the baseline dynamic factor model together with stochastic volatility for nowcasting the density of the GDP. For the baseline model, we closely follow Banbura et al. (2013) with the addition of the stochastic volatility. The key component of the model constitutes the incorporation of the survey expectations about the first and second moments of the predictive distribution of the GDP to the baseline model.

Econometric models of GDP nowcasting with mixed frequency datasets often involve GDP as the only quarterly variable that is complemented with a bulk of monthly or higher frequency variables that are timely available. We first demonstrate the model in terms of monthly variables in the set of high

<sup>2</sup>We provide full details about the simulation-based inference of the competing models in Section A of the supplementary material for the sake of brevity.

frequency variables and then we incorporate the GDP to the model.<sup>3</sup> Let  $y_t^m = [y_{1,t}^m, y_{2,t}^m, \dots, y_{n_m,t}^m]'$  for  $t = 1, 2, \dots, T$  denote the  $n_m$ -dimensional vector of variables in period  $t$ . The superscript of “ $m$ ” denotes the variables at monthly frequency. The variables are transformed to ensure stationarity if necessary and further standardized. We assume that these variables admit a factor structure as

$$y_t^m = \lambda^m f_t^m + \varepsilon_t^m, \quad \varepsilon_t^m \sim N(0, \Sigma), \quad (1)$$

where  $f_t^m$  is a vector of latent common factors,  $\lambda^m$  is  $n_m$ -dimensional vector of factor loadings and the covariance matrix, denoted as  $\Sigma$ , is of diagonal structure with the diagonal elements as  $\sigma_1^2, \sigma_2^2, \dots, \sigma_{n_m}^2$ . We assume that idiosyncratic factors, denoted as  $\varepsilon_t^m = [\varepsilon_{1,t}^m, \varepsilon_{2,t}^m, \dots, \varepsilon_{n_m,t}^m]'$ , are uncorrelated with  $f_t^m$  for all leads and lags.

For the dynamics of the common factor, we proceed with a single factor structure that obeys an AR(1) specification as

$$f_t^m = \phi f_{t-1}^m + u_t^m, \quad u_t^m \sim N(0, \sigma_u^2). \quad (2)$$

As noted in Banbura et al. (2013), employing more than one factor together with higher order autoregressive dynamics do not change the results qualitatively. Additionally, such simple specification keeps the model parsimonious and tractable for further extensions.

Since GDP is a quarterly flow variable, a special attention is required on its link to the monthly factor. Let  $z_t^q$  and  $z_t$  denotes the logarithm of quarterly and monthly (unobserved) GDP, respectively. Further, let  $y_t^q$  and  $y_t$  be the growth rates of the quarterly and monthly (unobserved) GDP, respectively. Following the approximation  $z_t^q \approx \sum_{s=0}^2 z_{t-s}$  employed by Mariano and Murasawa (2003) and Bańbura and Modugno (2014), among others,  $y_t^q$  is linked to  $y_t$  as

$$\begin{aligned} y_t^q &= \left( \sum_{s=0}^2 z_{t-s} \right) - \left( \sum_{s=0}^2 z_{t-3-s} \right) \\ &= \Delta z_t + 2\Delta z_{t-1} + 3\Delta z_{t-2} + 2\Delta z_{t-3} + \Delta z_{t-4} \\ &= y_t + 2y_{t-1} + 3y_{t-2} + 2y_{t-3} + y_{t-4} \\ &= \sum_{s=0}^4 w_s y_{t-s} \end{aligned} \quad (3)$$

where  $w_s$  denotes the weights for aggregation. Accordingly, we represent the measurement equation for the quarterly GDP growth rate as

$$y_t^q = \sum_{s=0}^4 w_s y_{t-s} = \sum_{s=0}^4 w_s (\lambda^m f_{t-s}^m + \varepsilon_{t-s}^m) = \lambda^q f_t^q + \varepsilon_t^q \quad (4)$$

if observed, otherwise treated as missing observation. Equation (4) implies an aggregation of the monthly common and idiosyncratic factors to the quarterly frequency where the quarterly factor  $f_t^q = \sum_{s=0}^4 w_s f_{t-s}^m$  for  $t = 3k$  and  $k = 1, 2, \dots, K$  where

$K$  is the number of quarterly observations. We assume that  $\varepsilon_t^q$  is white noise at quarterly frequency following Banbura et al. (2013). The final model can be cast as

$$\begin{aligned} y_t^q &= \lambda^q f_t^q + \varepsilon_t^q & \text{for } t = 3k \text{ and } k = 1, 2, \dots, K \\ y_t^m &= \lambda^m f_t^m + \varepsilon_t^m & \text{for } t = 1, 2, \dots, T \\ f_t^m &= \phi f_{t-1}^m + u_t & \end{aligned} \quad (5)$$

where  $\varepsilon_t^q$  and  $u_t$  are following Normal distributions with mean 0 and variances as  $\sigma_q^2$  and  $\sigma_u^2$ , respectively.  $\varepsilon_t^m$  follows a multivariate Normal distribution with means 0 and a diagonal covariance matrix  $\Sigma$  as discussed earlier. We provide the further details on the state-space representation of the general model in Section A of the supplementary material. The model provided in Equation (5) constitutes the main workhorse for nowcasting the GDP using the dynamic factor models. This model serves as the reference model in our performance evaluation.

Next, we incorporate stochastic volatility to the baseline model for nowcasting the density of GDP growth. For doing this, we allow for the variance of the GDP to evolve over time as

$$h_t = h_{t-3} + \eta_t, \quad \eta_t \sim N(0, \sigma_\eta^2) \quad (6)$$

where,  $h_t = \log(\sigma_{q,t}^2)$  and  $t = 3k$ ,  $k = 1, 2, \dots, K$  with the  $\sigma_{q,t}^2$  being the conditional variance of the GDP in time period  $t$ . Furthermore, we also incorporate a stochastic volatility process for the conditional variance of the factor as follows

$$h_{f,t} = h_{f,t-1} + \eta_{f,t}, \quad \eta_{f,t} \sim N(0, \sigma_{\eta_f}^2) \quad (7)$$

where,  $h_{f,t} = \log(\sigma_{u,t}^2)$  and for  $t = 1, 2, \dots, T$  with the  $\sigma_{u,t}^2$  being the conditional variance of the factor in time period  $t$ , see Marcellino, Porqueddu, and Venditti (2016) for a similar specification. Allowing for stochastic volatility in the factor structure enables us to capture any remaining time variation in the common volatility of the variables in the system.<sup>4</sup>

## 2.1. Incorporating the Density of Survey Expectations to the Baseline Statistical Model

We extend the baseline statistical models displayed in Equation (5) together with/without stochastic volatilities in Equations (6)–(7) with the information provided by the survey participants. We do this by using the first and the second moments derived from the expectations of the survey participants. We use Survey of Professional Forecasters (SPF) for the survey information about the quarterly GDP, which is released in the second month of the current quarter. The release dates of the SPF as well as the real GDP values are crucial for the model setup and the timing of the predictions. We perform nowcasts and forecasts at the end of each month recursively. As the GDP is released at the end of the first month following the quarter, this

<sup>3</sup>Banbura et al. (2013) show that variables that are sampled at higher frequency than monthly provide little or no improvement for prediction of the GDP.

<sup>4</sup>We also estimate models where monthly variables bear idiosyncratic stochastic volatility. This model provides less accurate predictions than the model with stochastic volatility structures in real GDP and factor error terms. The results of this specification search and the related discussion can be found in Section B of the supplementary material.



leads us to perform nowcasts rather than backcasts.<sup>5</sup> We provide a detailed timeline of release dates with the time of estimation represented in blue dots at the end of months in the following demonstration,

First, we start with the model that incorporates the first moment of these expectations to the baseline model without stochastic volatility. In the next part, we proceed to extend the baseline model with stochastic volatility further with the second moment of these expectations.

### 2.1.1. Incorporating the Mean

Let  $E_t^S[y_{t+h}^q]$  be the expectations of the survey participants of  $h$  period ahead GDP growth, where  $h = 3k+1$  for  $k = 0, 1, 2, 3, 4$ . Here  $S$  stands for the Survey implying that the expectation is computed using the predictions of the survey participants for the current quarter,  $k = 0$ , that is, nowcasts, for the next quarter,  $k = 1$ , until a year ahead,  $k = 4$ . That is,  $E_t^S[y_{t+h}^q]$  is available for the months  $t = 3k - 1$  for  $k = 1, 2, \dots, K$ . We illustrate the timeline of the survey data structure in the following figure,

Here SPF denotes the Survey of Professional Forecasters that is released in the second month of the first quarter of the reference year. The release is constituted by the expectations of the survey participants for the current quarter, which is represented by the short arrow stretched until  $t+1$ , expectations for the next quarter, represented by the arrow stretched until  $t+4$ , and so on. We incorporate the survey-based expectations of the GDP to the baseline statistical model by aligning the model based expectations with those derived from the survey. To do that, we first compute the expectations implied by the statistical model. Let  $E_t^M[y_{t+h}^q]$  be the  $h$ -period ahead GDP growth prediction of the model as described in Equation (5), where  $h = 3k+1$  for  $k = 0, 1, 2, 3, 4$ . Here  $M$  stands for the baseline statistical Model. In line with the timeline of the survey-based expectations, we first compute the model based expectations in the second month of the quarter,  $t$ , for the current quarter GDP growth,  $t+1$ , following (4) as

$$\begin{aligned} E_t^M[y_{t+1}^q] &= \lambda^q(E_t^M[f_{t+1}^m] + 2f_t^m + 3f_{t-1}^m + 2f_{t-2}^m + f_{t-3}^m) \\ &= \lambda^q(\phi f_t^m + 2f_t^m + 3f_{t-1}^m + 2f_{t-2}^m + f_{t-3}^m) \end{aligned} \quad (8)$$

where in the second line, the prediction of  $f_{t+1}^m$  is computed using the transition equation of the monthly factors that follows an AR(1) process. Notice that  $k = 0$  and thus  $h = 1$  corresponds to the current quarter real GDP prediction. Therefore, Equation (8) is a *nowcasting* equation. Similarly, for the *forecast* of the next quarter which corresponds to the time period  $t+4$ , the model based expectation can be computed as

$$\begin{aligned} E_t^M[y_{t+4}^q] &= \lambda^q(E_t^M[f_{t+4}^m] + 2E_t^M[f_{t+3}^m] + 3E_t^M[f_{t+2}^m] \\ &\quad + 2E_t^M[f_{t+1}^m] + f_t^m) \\ &= \lambda^q(\phi^4 f_t^m + 2\phi^3 f_t^m + 3\phi^2 f_t^m + 2\phi f_t^m + f_t^m). \end{aligned} \quad (9)$$

For the 1-quarter ahead forecasts, no lags of  $f_t^m$  are required as any values before  $t$  ( $t$  being the second month of a quarter) are too distant to enter the forecast for the future quarter. In general, for the forecast horizons  $h = 3k+1$  for  $k = 1, 2, 3, 4$ , that is, up to four quarter ahead, we can formulate the model based predictions as

$$\begin{aligned} E_t^M[y_{t+h}^q] &= \lambda^q(E_t^M[f_{t+h}^m] + 2E_t^M[f_{t+h-1}^m] + 3E_t^M[f_{t+h-2}^m] \\ &\quad + 2E_t^M[f_{t+h-3}^m] + E_t^M[f_{t+h-4}^m]) \\ &= \lambda^q \left( \sum_{s=0}^4 w_s \phi^{h-s} \right) f_t^m. \end{aligned} \quad (10)$$

The focus of the now/forecasting exercise is efficiently combining the predictions provided by the statistical model as described in Equations (8) and (10) with the predictions provided by the survey participants. We do that by aligning these two sources of predictions by extending the baseline statistical model in Equation (5) with further measurement equations following this aim. Specifically, we extend the model with the following measurements using the dataset provided by the surveys as

$$\begin{aligned} E_t^S[y_{t+h}^q] &= E_t^M[y_{t+h}^q] + \psi_{h,t} \\ \text{with } h &= 3k+1 \text{ for } k = 0, 1, 2, 3, 4, \end{aligned} \quad (11)$$

which leads to the following new measurement equations as nowcast and forecast equations

$$\begin{aligned} E_t^S[y_{t+1}^q] &= \lambda^q(\phi f_t^m + 2f_t^m + 3f_{t-1}^m + 2f_{t-2}^m + f_{t-3}^m) + \psi_{1,t} \\ E_t^S[y_{t+h}^q] &= \lambda^q \left( \sum_{s=0}^4 w_s \phi^{h-s} \right) f_t^m + \psi_{h,t} \\ \text{with } h &= 3k+1 \text{ for } k = 1, 2, 3, 4. \end{aligned} \quad (12)$$

Equation (12) implies that the statistical model based expectations should be in line with the expectations obtained from surveys subject to occasional differences or error terms,  $\psi_{h,t}$ , that follows a normal distribution with variance  $\sigma_{\psi_h}^2$ .<sup>6</sup> Equation (5) together with the measurement equations (12) constitute the final model where we combine the two sources of predictions in a statistically coherent way. For exploring the value-added provided by the survey information depending on the horizon of predictions, we estimate five models where we include, first, only a single measurement equation involving the survey nowcasts, that is,  $k = 0$ , and then we add the remaining measurement equations involving the survey forecasts successively for  $k = 1, 2, 3, 4$ .

<sup>5</sup>The timing for the release of survey-based predictions was also unclear before 1990. For the periods before the 2000s, survey-based predictions were released toward the end of the second month of the corresponding quarter, while after the 2000s it was released at the end of the second week rather than at the end of the month. Performing the predictions at the end of the month ensures that we do not use information that is not available in real-time. For the release dates of the survey information after the second quarter of 1990, see <https://www.philadelphiafed.org/-/media/frbp/assets/surveys-and-data/survey-of-professional-forecasters/spf-release-dates.txt?la=en&hash=B00319>

<sup>6</sup>We also consider stochastic volatility structure in these alignment equations. Results are unaffected when we allow for time variation in  $\sigma_{\psi_h}^2$ , and frequently it does perform worse than the models without stochastic volatility. The results of this specification and the related discussion can be found in Section B of the supplementary material.

### 2.1.2. Incorporating the Variance

While there is a consensus on the use of the surveys for predicting the real GDP growth, there is an intense debate on the accurate measurement of uncertainty surrounding these predictions derived from the survey information, see, for example, Boero, Smith, and Wallis (2008), Rich and Tracy (2010), Abel et al. (2016), and Rich and Tracy (2021) among others. When the focus is on using individual point predictions of survey participants as aggregate density functions, as in our case, **two measures of uncertainty come forefront. These include, first, the “disagreement” measured as the variance of the predictions provided by the participants in a given period; see, for example, Bomberger (1996) for an earlier analysis. Second, the ex-post measure of uncertainty (EPU), which considers the variance of the prediction errors over time obtained by the difference between the actual realization of real GDP growth and average point prediction from surveys, see the discussion in Clements (2014), and the practice in Aastveit et al. (2014), and Tallman and Zaman (2019).** On the one hand, the disagreement provides noisy but timely measures of uncertainty the forecasters face in a given period. On the other hand, the EPU provides an accurate measure of uncertainty but uses limited data due to delay in the realization of the real GDP to measure the prediction error. Our specification search on these measures of uncertainty derived from the survey-based predictions indicates that both measures perform similarly in our model framework with a slightly better performance of the disagreement measure. Therefore, we opt to use the disagreement measure for the second moment of survey-based predictions.<sup>7</sup> Let  $D_{t,h}$  denote the disagreement among the survey participants in period  $t$  about the  $h$ -period ahead GDP growth. This can be computed as

$$D_{t,h} = \frac{1}{N_t} \sum_{i=0}^{N_t} \left( S_{i,t+h} - E_t^S[y_{t+h}^q] \right)^2 \quad (13)$$

where  $N_t$  is the number of survey participants, and  $S_{i,t+h}$  is the  $h$ -step ahead prediction of participant  $i$  in period  $t$ . We align the disagreement computed in each quarter for the  $h$ -period ahead predictions of the survey participants with the predictions of volatility from the baseline statistical model with stochastic volatility derived using Equation (6). Specifically, assuming a random walk process for the (log-)volatility,  $h$ -period ahead predictions,  $E_t^M[h_{t+h}]$ , correspond to the current volatility,  $h_t$ . Therefore, we incorporate the disagreement,  $D_{t,h}$ , computed using the survey information, with the model based predictions as follows

$$\log(D_{t,h}) = h_t + \xi_{t,h} \quad \text{with } h = 3k + 1 \text{ for } k = 0, 1, 2, 3, 4. \quad (14)$$

This implies five additional measurement equations that serve for the combination of the (log-) variance obtained from the baseline statistical model together with the (log-) disagreement obtained from surveys. As in the case of the first moment, we estimate five models where we include, first, only a single measurement equation involving the disagreement among survey nowcasts, that is,  $k = 0$ , and then we add the remaining measurement equations involving the disagreement among survey forecasts one by one for  $k = 1, 2, 3, 4$ .

<sup>7</sup>We provide our findings on comparing the two measures in Section C of the supplementary material.

## 3. Data and Estimation Details

### 3.1. Data

We consider the U.S. real GDP over the period starting from the last quarter of 1968 until the end of 2019 for the measure of output.<sup>8</sup> This analysis is intended for performance evaluation of competing models without including the extreme periods of the Covid-19 pandemic in 2020. We provide a detailed month-by-month analysis in the following sections for the so-called “pandemic recession” periods and the drastic bounce-back after that. We construct a broad set of monthly variables involving 16 variables for the monthly dataset.<sup>9</sup> These variables are employed in Banbura et al. (2013) as well, and thus, it provides ample opportunities to compare the resulting model with the existing popular and successful models, which we use as the benchmark.<sup>10</sup> For the survey data, we use the predictions from the Survey of Professional Forecasters (SPF) published by the Federal Reserve Bank of Philadelphia.<sup>11</sup> We provide details on these variables in Table 1.

In particular, we consider the predictions of the survey participants for the real GDP growth at the individual level. We display the evolution of the mean of these predictions and the disagreement among the forecasters as computed in Equation (13) in left and right panel of Figure 1.

The left panel of Figure 1 shows that the mean of the survey predictions tracks the real GDP growth smoothly with limited variation, especially during expansions. This variation further reduces with longer horizon predictions. The decline in the current real GDP growth predictions coincides quite accurately with the actual recession dates, displayed with the gray shaded areas. Note that while these predictions are provided in a timely manner, actual real GDP values are released with a lag.

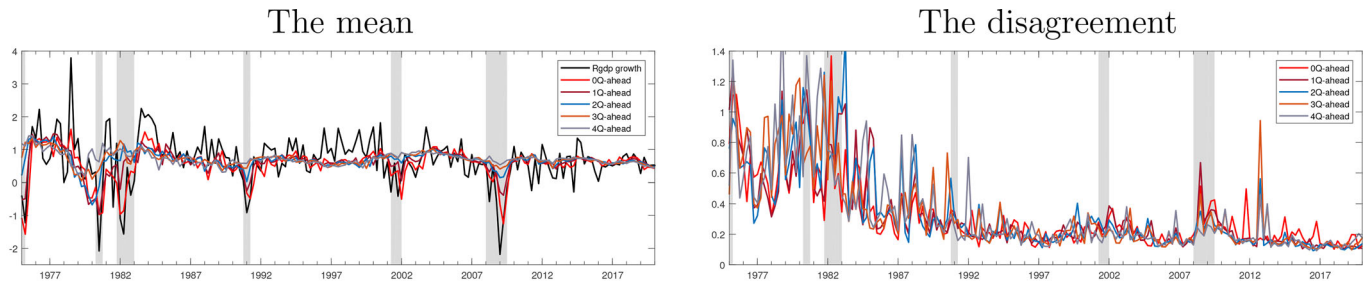
Considering the right panel of Figure 1, the disagreement before the 1980s lessens considerably after the mid-1980s in line with the notion of great moderation, which refers to the decline of variation in many U.S. macroeconomic series, see McConnell and Perez-Quiros (2000), among others, for details. We observe that the periods before the mid-1980s are comprised of quite erratic swings around high levels of disagreement. On the contrary, the uncertainty is typically relatively low after the

<sup>8</sup>Essentially, the output is measured as Gross National Product (GNP) until 1991 and Gross Domestic Product (GDP) after that. Still, the output is denoted as GDP for the whole sample period for the sake of demonstration.

<sup>9</sup>As an alternative, we consider a much larger dataset using 135 variables obtained from the FRED-MD dataset of McCracken and Ng (2016). These models perform inferiorly compared to those using the dataset involving 16 variables. We provide details on this comparison in Section C of the supplementary material.

<sup>10</sup>The original dataset of Banbura et al. (2013) includes some variables at daily and weekly frequency as well, including return on the market index at the daily frequency and initial jobless claims at the weekly frequency. They note that the daily and weekly variables do not provide further information beyond the monthly variables for nowcasting GDP. We follow this practice to exclude the variables at higher than monthly frequency. This structure also facilitates the computation substantially for the recursive out-of-sample exercise without deteriorating the model's overall performance. We also consider alternative model strategies where the survey information is used as typical data as in Equation (5) with using the alignment equations as in Equation (12). These models perform inferiorly compared to our framework. We provide details on this comparison in Section B of the supplementary material.

<sup>11</sup><https://www.philadelphiafed.org/research-and-data/real-time-center/survey-of-professional-forecasters>



Note: The left panel displays the evolution of the mean of the SPF predictions, and the right panel displays the disagreement among the forecasters.

**Figure 1.** The mean and the disagreement of SPF based predictions.

NOTE: The left panel displays the evolution of the mean of the SPF predictions, and the right panel displays the disagreement among the forecasters.

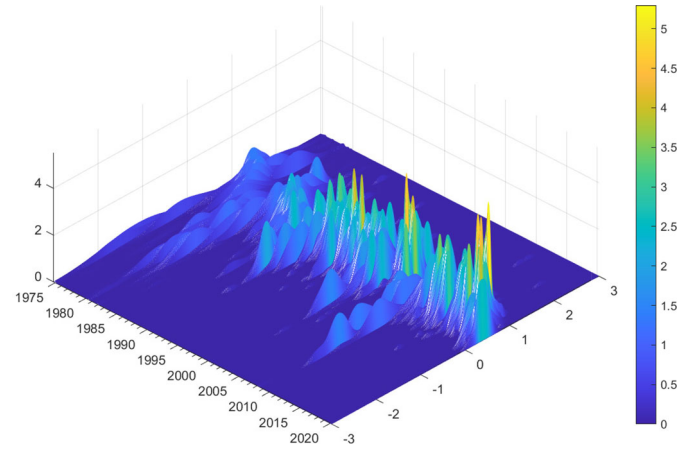
**Table 1.** Dataset: Frequencies, transformations, periods and publication delays.

No	Series	F	T	Start	Delay
1	Real Gross Domestic Product	Q	1	1968:IV	1M
2	Industrial Production Index	M	1	1968:12	1M
3	Purchasing Manager Index: Manufacturing	M	2	1968:12	1M
4	Real Disposable Personal Income	M	1	1968:12	1M
5	Unemployment Rate	M	2	1970: 1	1M
6	All Employees: Total Nonfarm Payrolls	M	1	1968:12	1M
7	Personal Consumption Expenditures	M	1	1968:12	1M
8	Housing Starts: Total: New Privately Owned Housing	M	1	1968:12	1M
9	New One Family Houses Sold	M	1	1968:12	1M
10	Manufacturers' New Orders: Durable Goods	M	1	1992: 2	1M
11	Producer Price Index: Finished Goods	M	1	1968:12	1M
12	Consumer Price Index for All Urban Consumers	M	1	1968:12	1M
13	Imports of Goods and Services	M	1	1992: 1	2M
14	Exports of Goods and Services	M	1	1992: 1	2M
15	Philadelphia Fed Survey: General Business Conditions	M	2	1968:12	–
16	Retail Sales: Retail and Food Services	M	1	1992: 1	1M
17	Conference Board Consumer Confidence	M	2	1968:12	–
	Survey of Professional Forecasters				
	Real Gross Domestic Product	Q	1	1968:IV	-1M

NOTE: T indicates the type of transformation of variables to ensure stationarity (1 = first difference of logarithm, 2 = first difference) and F indicates frequency. Series at higher frequencies are converted to monthly frequency by using corresponding frequency averages. Our reference point for delays is the end of months in which we produce the nowcasts. The variable highlighted with the dark gray is the data obtained from the predictions of the Survey of Professional Forecasters for real GDP. “-1M” indicates that predictions up to four quarters ahead are available in the second month of the current quarter.

mid-1980s but is aggravated almost instantly during turbulent periods that rapidly reverses back afterward. The disagreement in current quarter predictions, that is, survey-based nowcasts, is typically larger than the corresponding forecasts, as seen in many periods of severe recessions of 1982 and 2007. This difference is because long-horizon predictions tend to follow the long-run level, which produces less variation than current values predictions. As the forecasts are provided  $h$ -periods earlier than the realizations, and the nowcasts are provided concurrently, a shock to the economy has an instant effect on nowcasts, which translates into more considerable disagreement among these predictions. For a visual representation of survey-based nowcasts together with the implied mean and uncertainty, that is, predictive density for the current quarter GDP growth, we display (kernel approximations of) the distributions over time in Figure 2.

Rapid changes in the location of the distributions during recessions together with the changing uncertainty can nicely be traced in Figure 2. The light color indicates the thinner



**Figure 2.** The distribution of the SPF based predictions of current quarter real GDP growth.

NOTE: The figure displays the evolution of the distribution of the SPF-based predictions for the current quarter real GDP growth from 1977 until 2019.

distributions, representing a limited amount of uncertainty over the 1990s, while the darker colors imply broader distributions indicating the sudden changes in uncertainty during severe recessions and the periods before the mid-1980s.

We first conduct a full sample estimation for the analysis of the models. Next, we perform a recursive out-of-sample analysis in pseudo-real-time using the ragged edge datasets due to publication delays to evaluate the model performance with/out survey information.

### 3.2. Estimation Procedure and Evaluation of Models

We use Bayesian inference using Markov chain Monte Carlo (MCMC) simulation techniques for estimation and inference in the unobserved component model and its variants. Specifically, we use Metropolis within Gibbs sampling together with data augmentation (see Tanner and Wong 1987) to obtain posterior results. While the baseline statistical model allows for using plain MCMC involving only Gibbs sampling using standard conditional distributions, the fact that incorporating survey information leads to nonlinear parameter structure as in Equation (12). We employ the Metropolis-Hastings algorithm in such cases.<sup>12</sup> The posterior distribution for the model parameters

<sup>12</sup>The stochastic volatility structures in our model framework also lead to nonlinearities as the dependence structure of the log-volatilities to the data



is proportional to the product of the likelihood function and the prior distributions for model parameters. While for the likelihood specification we make use of the multivariate Normal distribution following the convention, for the prior specifications of the model parameters, we specify uninformative priors. Details on the prior specifications and the resulting posterior distributions along with the simulation scheme are provided in Section A of the supplementary material.

We compare the competing models with and without survey information to uncover the value-added provided by the survey information on top of the baseline statistical models. We use the marginal likelihood metric to explore the performance of models using the full sample. For the out-of-sample density and point prediction comparisons, we use the conventional measures of predictive likelihood and Root Mean Squared Forecast Error (RMSFE) metrics.<sup>13</sup> While the data spans the periods over the last month (last quarter) of 1968 until the end of 2019, we consider the period starting from the first month (first quarter) of 1977 until the end of the sample as the evaluation period to compute the predictive likelihoods.

#### 4. Empirical Findings

In this section, we discuss our empirical findings based on the estimation results of various models. We first estimate the baseline statistical model as described in Equation (5), denoted as “BM” referring to the Baseline Model. We extend this model using the mean predictions obtained by SPF each time adding  $k$ -quarter ahead predictions for  $k = 0, 1, \dots, 4$  as described in Equation (11). We denote these models as “(BM+S)-M $k$ ” for  $k = 0, 1, 2, 3, 4$  referring to Baseline Model and Survey information with Mean aligned using up to  $k$ th-quarter ahead predictions. Comparison of the BM with those extended using the first moment of survey-based predictive distributions would indicate, first, whether survey-based predictions of mean bear additional information and second, whether this information is embedded only in survey-based nowcasts or it is also carried over in forecasts at various horizons.

Next, we extend the BM with the stochastic volatility using the specification described in Equations (6) and (7). This extension implies conducting a thorough density estimation with time variation in both the first and second moments. This model is denoted as “BMSV.” We extend this model using the mean predictions obtained by SPF as in the previous case. We denote these models as “(BMSV+S)-M $k$ ” for  $k = 0, 1, 2, 3, 4$ . Finally, we further extend this model using the measure of disagreements among the survey participants obtained by SPF as described in Equation (14). This alignment implies that the last group of models exploits both the first and second moments of the predictive distributions from SPF, essentially combining these with the predictive distributions obtained from the baseline statistical model. We denote these models as “(BMSV+S)-MV $k$ ” for  $k = 0, 1, 2, 3, 4$ .

#### 4.1. Full Sample Findings

We start by evaluating the main estimation results of selected competing models. For visual inspection of the main findings in conforming stylized facts, Figure 3 displays the predicted mean, that is, the fitted values obtained using the BMSV model.

As can be seen from Figure 3, the model estimated using the full sample data provides accurate in-sample predictions and it performs very well in tracking the mean real GDP growth rate throughout the sample periods. The timing of recessions as well as recovery periods are captured by the statistical model successfully.<sup>14</sup>

Figure 4 displays the predicted volatility using the BMSV model. We also include the estimates obtained from the (BMSV+S)-MV4 model that combines both the mean and the disagreements obtained from the SPF together with the BMSV model for comparison.

The estimation results imply that a rapid decline follows the 1970s and early 1980s volatile periods in the mid-1980s consistent with the great moderation. We observe increases in volatility during recessions in the 2000s, albeit much limited compared to the levels of volatility before great moderation. A critical aspect of the volatility estimated by the BMSV model is that it evolves smoothly over time. On the contrary, volatility estimates obtained by the (BMSV+S)-MV4 model involve more variation. It seems that (BMSV+S)-MV4 model provides a compromise between the smoothly evolving volatility obtained by the BMSV model and the rapidly changing nature of ambiguity captured by the disagreement of participants of SPF. Figures 3 and 4 indicate that the main competing models can capture the moments of the real GDP growth quite successfully.

Next, we consider the model fit based on the marginal likelihood metric computed using the full sample for all competing models. For the ease of model comparison, we provide the (log-)marginal likelihood values for the baseline statistical model (BM), and for the remaining models, we display the differences between the (log-)marginal likelihood of the competing models with that of the BM. These correspond to the Bayes factors of competing models with respect to BM. We display these in the first three columns of Table 2. Different groups of models are given in rows with varying shades of gray for the ease of demonstration.

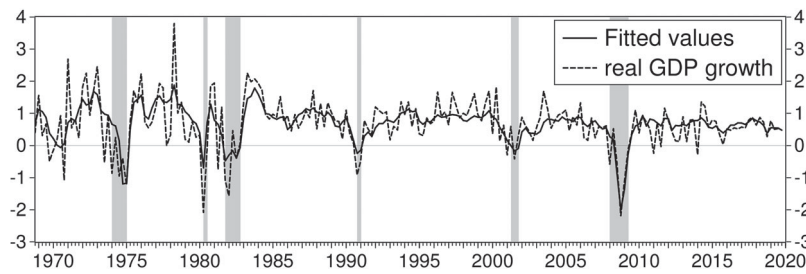
When we consider the baseline model in the first row, it is seen that the data flow does not matter much. The (log-)marginal likelihood values are around  $-226$  regardless of the month of the quarter the predictions are performed as can be seen in the first row of Table 2. When these predictions are aligned with the first moment of survey-based predictions displayed in the next panel with darker gray shade, we observe that survey-based nowcasts, that is, predictions of the current quarter by the survey participants, do not cause any improvement in the marginal likelihoods. However, this picture reverses when we incorporate survey forecasts, that is, predictions of next quarters by the survey participants. In this case, marginal

is through the exponential function. In these cases, we use the approach of Omori et al. (2007) where we approximate the model using a mixture of Normals. We provide details on the estimation of volatility in the Section A of the supplementary material.

<sup>13</sup>Details on the computations of these measures are provided in Section A of the supplementary material.

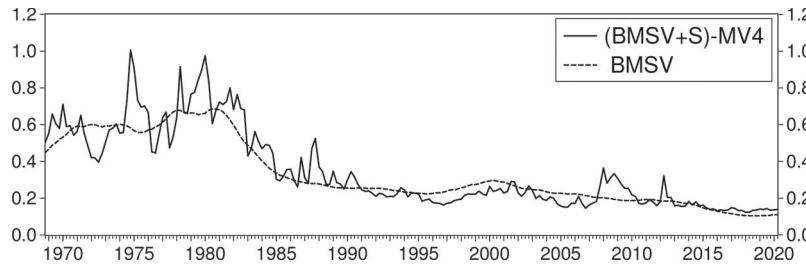
<sup>14</sup>Here we do not display the fitted values obtained from (BMSV+S) type of models for the clarity of the demonstration as these are very similar to those displayed in Figure 3.





**Figure 3.** Fitted values of GDP growth obtained using the BMSV model.

NOTE: The figure displays the evolution of the fitted values of real GDP growth obtained using the BMSV model estimated using the full sample period from 1977 until 2019.



**Figure 4.** Comparison of volatility estimates.

NOTE: The figure displays the evolution of the volatility estimates obtained using the BMSV and (BMSV+S)-MV4 models estimated using the full sample period from 1977 until 2019.

**Table 2.** The marginal and predictive likelihoods of competing models.

	Marginal likelihoods			Predictive likelihoods								
	Q0			Q0			Q1			Q2		
	M1	M2	M3	M1	M2	M3	M1	M2	M3	M1	M2	M3
BM	-226.8	-225.8	-226.3	-176.5	-162.2	-154.6	-196.0	-193.9	-187.1	-200.9	-199.6	-198.1
(BM+S)-M0	0.1	0.8	0.4	4.4	3.1	1.2	5.7	4.6	4.4	5.5	4.7	4.8
(BM+S)-M1	7.0	6.6	5.8	5.1	6.4	3.2	5.8	7.2	5.9	5.1	5.5	4.7
(BM+S)-M2	7.8	9.2	9.1	5.9	6.0	3.5	5.4	7.0	5.5	4.4	4.4	4.8
(BM+S)-M3	8.2	8.5	9.7	5.3	4.9	2.2	6.1	7.7	5.8	3.7	4.3	4.4
(BM+S)-M4	9.5	9.5	9.0	6.1	4.7	2.7	5.1	7.4	6.0	2.7	5.3	4.0
BMSV	0.1	-0.4	-2.1	16.5	17.8	17.7	17.9	18.7	18.9	17.2	17.6	17.8
(BMSV+S)-M0	1.5	0.2	0.4	20.1	18.0	17.2	21.3	21.0	21.9	21.7	22.7	22.4
(BMSV+S)-M1	8.4	7.2	6.4	21.3	20.1	18.9	23.4	23.5	21.7	21.8	22.9	22.1
(BMSV+S)-M2	8.9	11.3	9.2	21.2	22.3	20.3	22.0	24.2	21.4	20.6	21.1	22.9
(BMSV+S)-M3	11.5	7.8	11.4	22.1	21.2	21.0	22.8	21.1	22.3	22.9	21.1	22.0
(BMSV+S)-M4	12.7	5.3	9.5	21.0	20.8	20.1	21.0	22.9	21.4	19.6	22.1	21.0
(BMSV+S)-MV0	3.8	9.0	10.9	24.2	23.5	22.3	24.3	22.3	22.7	23.7	21.4	23.7
(BMSV+S)-MV1	15.2	16.0	17.6	25.5	26.0	24.1	25.5	24.3	23.7	22.6	25.3	23.0
(BMSV+S)-MV2	14.9	13.5	19.8	25.8	25.0	24.4	23.9	25.3	24.8	22.6	23.2	25.0
(BMSV+S)-MV3	19.7	14.8	17.7	26.3	26.1	24.5	25.5	25.9	24.5	25.3	25.1	24.2
(BMSV+S)-MV4	20.5	18.2	17.1	26.2	26.6	25.9	25.6	26.5	25.4	22.1	24.9	24.0

NOTE: The marginal likelihoods are calculated for the first, second, and third months of the current quarter for the period starting from 1968 until 2019. The predictive likelihoods are computed at the end of each month in the current quarter, for the current quarter (Q0-M1, Q0-M2, Q0-M3), which implies nowcasts, for one quarter ahead (Q1-M1, Q1-M2, Q1-M3) and for two quarters ahead (Q2-M1, Q2-M2, Q2-M3), which imply forecasts, over the evaluation period starting from 1977 until 2019 recursively. BM stands for the baseline model as described in Equation (5) and (BM+S)-M<sub>k</sub> stands for the model where we extend the BM model with the first moment of the survey-based predictions for  $k = 0, 1, \dots, 4$  as described in Equation (11). BMSV stands for the baseline model together with stochastic volatility as described in Equations (6) and (7). Finally, (BMSV+S)-MV<sub>k</sub> stands for the model where we extend the BMSV model with the first and the second moment of the survey-based predictions for  $k = 0, 1, \dots, 4$  as described in Equations (12) and (14). Statistical significance of the predictive Bayes factors are tested using the Diebold–Mariano (DM) test using the measures of predictive likelihood contributions together with the HAC covariance matrix and a finite sample correction, Harvey, Leybourne, and Newbold (1997). The cells with white background contain the values that are statistically *insignificant* at the conventional significance level of 5%.

likelihood values increase by around 9 points. Thus, the mean of survey-based predictions of the future periods possesses additional information.

When the two baseline statistical models with and without stochastic volatility are compared, allowing for time variation in volatility deteriorates model fit marginally, as seen from the mostly negative Bayes factor values corresponding to the BMSV

model. When this model is extended using the alignment with the mean of survey-based predictions of future periods, we observe positive Bayes factors by around 10 points. These differences show that the survey participants' predictions of future periods rather than the current period bear indeed additional information beyond that is contained in the extracted factor and/or volatility.

The largest improvement is obtained when survey-based predictions are incorporated for both the first and second moments. The last panel with the darkest shade of gray indicates that the disagreements among the survey participants on future periods' expectations matter, but the disagreement on nowcasts also significantly improves marginal likelihood values. When the first and second moments of the predictive distributions implied by the baseline model are aligned with those of the survey-based nowcast distribution as it is the case for the (BMSV+S)-MV0 model, we observe an increase around 9 points over the conventional baseline statistical model and an increase around 10 points over the baseline statistical model with stochastic volatility. When we incorporate further the future predictions based on the survey both in terms of the predicted mean and the disagreement among the survey participants, marginal likelihood values increase by 20 points. These results decisively indicate that survey-based predictions of the full density of the real GDP growth embed pivotal information over the conventional baseline statistical models.

#### 4.2. Predictive Performance

While the marginal likelihood based evaluation of the models relies on the dataset over the full sample period, it is crucial to examine the performance of the models in real-time using the information available at the time of prediction. Therefore, in this section, we evaluate the predictive performance of the models based on predictive likelihoods. For testing the statistical significance of the (log-)predictive likelihood differences of competing models compared to the baseline statistical model BM, we compute the Diebold-Mariano (DM) test using the predictive likelihood contributions together with HAC covariance matrix and a finite sample correction as discussed in Harvey, Leybourne, and Newbold (1997).

Predictive likelihood values are displayed in Table 2 starting from the fourth column. The fourth to sixth columns under the column header "Q0" present the predictive likelihood values computed using nowcasts performed at the end of the current quarter's first, second, and third months. When we focus on the baseline statistical model, the effects of data flow can explicitly be seen in real-time, as the increase in predictive likelihoods over the quarter is as high as 22 points. Moreover, this increase in the predictive likelihood values can be seen for all cases independent of whether we allow for time variation in volatility and whether survey information is added or not.

Incorporating the mean nowcasts from the survey predictions adds on top of the baseline statistical model with increases in predictive likelihoods of around 3 points for the first two months of the quarter, as can be seen from the row corresponding to the (BM+S)-M0 model. However, the improvements are statistically insignificant at conventional significance levels except for the nowcasts performed at the end of the quarter's first month. The impact of the survey-based predictions on the baseline model increases further when we incorporate survey-based forecasts with an increase of up to 6 points. Except for some nowcasts performed at the end of the quarter, these improvements are all statistically significant. This significance is particularly the case in the second month of the current quarter when the survey information is released. Hence, incorporating survey-based forecasts of mean in addition to nowcasts of it

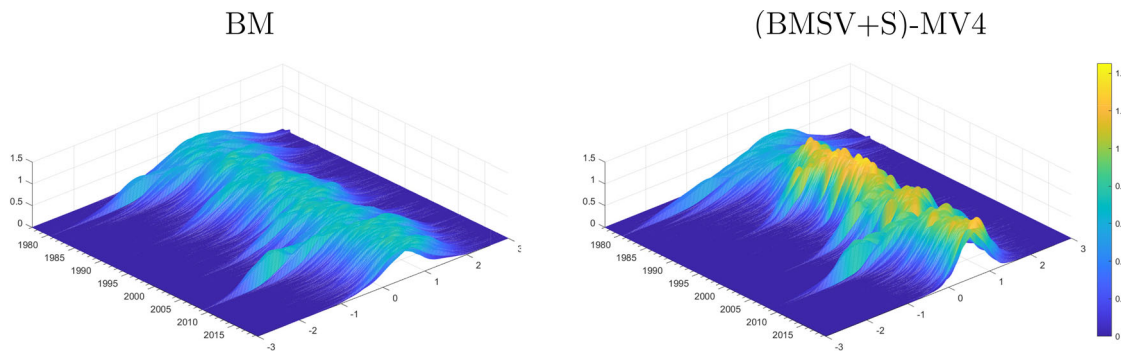
enhances the improvement of the predictive capability of the baseline model further.

When evaluating full sample results in the previous section, we see that incorporating stochastic volatility to the baseline model deteriorates the model fit. However, this picture reverses when we evaluate models in real-time using predictive likelihoods. Allowing for stochastic volatility in the baseline statistical model improves predictive likelihood by around 17 points when the row corresponding to BMSV is compared to BM. This substantial improvement is statistically significant as well. Furthermore, when the survey-based nowcasts of the mean are incorporated, we observe an improvement of 3 points for the nowcasts performed at the end of the first month of the quarter, which is statistically significant. Once again, this rises to around 5 points when incorporating survey-based forecasts on the first moment of the predictive distribution of the real GDP growth.

The impact of the survey information on the predictive ability of the baseline statistical model soars when we also incorporate the disagreement among the survey participants as a measure of the second moment, that is, volatility of the predictive distribution. In this case, incorporating only the nowcasts of survey-based predictive distribution, including the first and the second moment, increases nowcasts' predictive likelihoods by around 6 points. Furthermore, when survey-based predictive distributions using the survey participants' forecasts are incorporated, in addition to nowcasts of them, this gain increases to 9 points compared to the baseline statistical model with stochastic volatility without survey information.

To summarize, when the (BMSV+S)-MV4 model, where we incorporate all available survey information on the entire distribution of the real GDP, is compared to the plain BM model, the improvement in predictive likelihoods are as high as 27 points, which is statistically significant as well. For a visual representation of the real-time nowcasts obtained by the BM and (BMSV+S)-MV4 models, we display (kernel approximations of) the distributions, that is, the predictive densities for the current quarter real GDP growth, obtained from these models over time in Figure 5. We display these densities for the BM model in the left panel and the (BMSV+S)-MV4 model in the right panel of Figure 5. Compared to the SPF-based distribution provided in Figure 2, predictions offered by (BMSV+S)-MV4 model provide more smoothly evolving predictive distributions with relatively broader distributions during expansions. These distributions become thinner when we consider recessions mitigating the excessive uncertainty in the predictions of SPF. Compared to the BM model, predictive distributions offered by (BMSV+S)-MV4 model have lighter colors, indicating relatively smaller variance. Therefore, they are gathered densely around the mean of the nowcasts. These dense distributions improve the predictive likelihood values considerably as the uncertainty around the real GDP growth is further resolved when survey nowcasts are combined with the baseline statistical model, especially when the mean predictions of both models are similar. These findings suggest that survey-based predictions deliver significant and genuine information not contained in comprehensive datasets like ours, including many forward-looking variables such as PMI and confidence indices beyond conventional indicators.

Next, we evaluate the forecasting performance of the model by focusing on 1- and 2-quarter ahead forecasts in the last six columns of Table 2. Considering the impact of the data flow on the 1-quarter ahead forecasts, we observe a statistically signifi-



**Figure 5.** The distribution of the selected model-based nowcasts for the real GDP growth.

NOTE: The figure displays the evolution of the predictive nowcast distribution for the current quarter real GDP growth obtained using the BM and (BMSV-S)-MV4 models over the evaluation period from 1977 until 2019.

cant, sizable effect. The predictive likelihood values improve by around 9 points when predictions are performed in the third month of the current quarter, M3, compared to predictions performed in the first month of the current quarter, M1. However, this impact of the data flow vanishes when we consider 2-quarter ahead forecasts. The differences in predictive likelihood values drop to around 3 statistically insignificant points in this case. Therefore, we conclude that, while the impact of the data flow over the current quarter is largest when we consider nowcasting, this impact fades smoothly out with the increasing forecast horizon.

Regarding the impact of the survey-based information on forecasting, we see that in almost all cases, the information in survey nowcasts is sufficient to improve the predictive ability of the statistical model. In terms of the first moment, survey-based forecasts add predictive gains on top of the gains obtained by survey-based nowcasts. For example, enhancing the baseline statistical model, “BM,” by incorporating the mean extracted from survey-based nowcasts (BM+S)-M0 escalates the predictive likelihood value around 5 points, which is statistically significant at conventional significance levels as well. However, incorporating the first moments from survey-based forecasts, as is the case for the model (BM+S)-M4, improves the likelihood value around 3 points for 1-quarter ahead predictions (performed in the second month of the current quarter) which is significant at 5% significance level. Furthermore, as in the case of nowcasting, incorporating stochastic volatility to the baseline model, “BMSV,” enhances the predictive ability of the statistical model for both 1- and 2-quarter ahead forecasts. More importantly, aligning the first moment of the predictive distribution extracted from survey nowcasts and forecasts with that of the baseline statistical model with stochastic volatility increases the predictive likelihood by almost 5 points in this case. Furthermore, aligning the second moment from the survey information further improves predictive capability by an additional 3 points on average, significant at a 5% significance level in most cases.

### 4.3. Evolution of Predictive Gains from Survey Information Over Time

The findings displayed in the previous section show that survey enhanced model predictions possess useful information enhancing the predictive ability of the baseline statistical model both in terms of nowcasts and forecasts. In this section, we focus

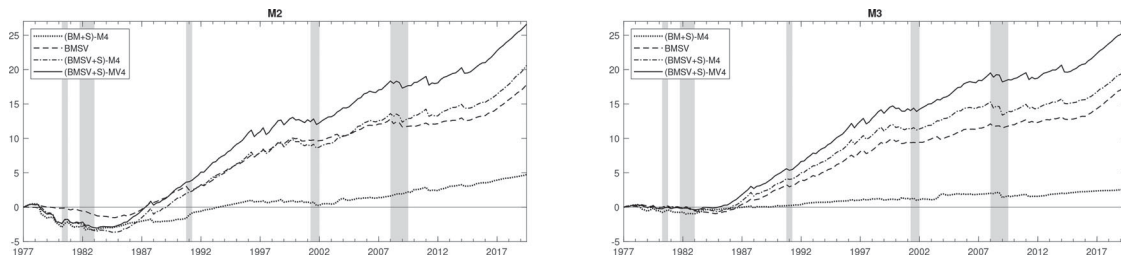
on the evolution of these predictive gains to explore further whether survey information has a particular pattern over time in improving the predictive capability of the baseline model. Figure 6 displays the evolution of the Bayes factors, which are computed as the differences of predictive likelihoods (computed recursively) between the competing models and the BM model. The predictive likelihoods are computed using the predictions performed in the second and third months of the current quarter, that is, using nowcasts of the current quarter.<sup>15</sup> Moreover, for the models that use survey information, we only include those models that incorporate all available survey information, including nowcasts and up to 4-quarter ahead forecasts.

Before analyzing the contribution of survey information throughout the evaluation period, we consider the evolution of the Bayes factor for the selected models at the onset of our evaluation sample until the mid-1980s. During these periods, all models but the BMSV model perform worse than the benchmark of the BM model. On the contrary, the outperformance of the survey enhanced models prevails from the mid-1980s until the end of the sample.<sup>16</sup> Nevertheless, this deterioration rapidly vanishes for the nowcasts of the third month of the quarter with the extension of the dataset with further data releases.

Regarding the impact of the survey information, we start to evaluate the dynamics of the value-added of incorporating the survey-based mean predictions to the statistical model. Therefore, we focus on the evolution of the Bayes factor for the (BM+S)-M4, relative to the plain baseline model, BM, as shown using the dotted line in Figure 6 for the nowcasts. We see that the Bayes factor evolves smoothly, favoring the (BM+S)-M4, almost like a straight line. We can trace some sudden changes

<sup>15</sup>We display the graph of the evolution of predictive likelihoods for predictions performed in the first month of the current quarter in Section E of the supplementary material as it is very similar to Figure 6.

<sup>16</sup>We conduct a detailed analysis for exploring the underlying reasons for this inferior performance at the onset of the evaluation sample. Results indicate that this limited performance is partly due to the inferior performance of surveys predictors during the 1970s, where we observe extreme turmoils. Combined with the lack of data at the sample's onset, this poor performance led to considerable deterioration of the predictions toward the end-1970s. Interestingly, this substantially poor performance of the surveys seems unique to this period. Essentially, we do not observe such performance of survey participants for latter turmoil periods. On the contrary, as discussed in Section 5 survey participants react pretty promptly to the changing conditions during the Covid-19 pandemic induced recession of 2020. We display those results in Section E of the supplementary material.



**Figure 6.** The evolution of Bayes factors over time computed using predictions from the selected models obtained at the end of the second and third months of the quarters.

NOTE: The figure displays the evolution of the Bayes factors for the (BM+S)-M4, BMSV, (BMSV+S)-M4, and (BMSV+S)-MV4 models relative to BM based on the predictions obtained at the end of the second and third months of the quarters over the evaluation period from 1977 until 2019.

in Bayes factors, including an increase around the recession of 1990–1991 and 2003 for the nowcasts performed in the second month. The increase in the Bayes factor in 1990–1991 is related to the timely prediction of the short-lasting mild recession by the survey participants relative to the statistical model. The sudden increase around 2003 corresponds to the start of a relatively faster growth path of U.S. real GDP, which seems to be anticipated by the survey participants swiftly. Therefore, while survey-based real GDP predictions are generally aligned with the statistical model, these provide valuable information beyond what is contained in the model during periods with abrupt changes.

Considering the impact of time variation in volatility, we include the evolution of the Bayes factor of the baseline model with stochastic volatility, BMSV, relative to the plain baseline model, BM, displayed using the dashed line. This impact can be seen after the mid-1980s with the start of great moderation when the variation in real GDP growth rates has reduced considerably. While the BMSV model accommodates this reduction throughout the 1980s and 1990s, the Bayes factors are dampened in the 2000s. Indeed, although the Bayes factor was around 10 in 2000, it increased only to 14 toward the end of the sample. Moreover, a large part of this increase took place after 2017. When we focus on the model where the (mean) survey predictions of the real GDP growth are incorporated into the BMSV model, that is, (BMSV+S)-M4, we observe that it performs constantly better than the BMSV model starting from the mid-1980s until the end of the sample.

Finally, when we incorporate the full predictive density in terms of the first and second moments from the survey information with the statistical model as is the case for the (BMSV+S)-MV4 represented using the solid line, we see that this model is dominant over all competing models throughout the evaluation period. We can track rapid and long-lasting increases throughout the evaluation period interrupted with rapid but short-lasting decreases mostly coinciding with depressive periods such as 2008, 2011, and 2013. These periods correspond to the collapse of Lehman Brothers, the European debt crisis, and the taper tantrum, respectively. Incorporating the second moment of predictive density extracted from the survey information leads to more visible and potent swings in Bayes factors. Hence, the impact on predictive performance due to survey information follows a sentimental wave. During turmoil periods, excessively increasing volatility rapidly deteriorates the “density” prediction, which lasts very short. On the contrary, this deterioration is followed by a long-lasting improvement

after a swift adjustment during good times following turmoil periods. Therefore, it considerably improves the overall performance of survey information augmented statistical models compared to the plain statistical model.

#### 4.4. Point Forecast Accuracy of Competing Models

The previous section provides compelling evidence on the survey augmented models’ predictive gains compared to the baseline statistical model. Results indicate that the first and second moments obtained from the surveys contribute significantly to the density prediction when they are aligned with those moments from the baseline statistical model. However, it would also be informative to know whether the survey information improves point forecasts. This section compares competing models for point predictive performance using the Root Mean Squared Forecast Error (RMSFE) metric. We first display the RMSFE of the competing models relative to RMSFE of the BM model in Table 3. We provide the raw RMSFE values for the baseline statistical model (BM) for the ease of model comparison. We also display relative RMSFE from the survey as these can be computed in this case.

The point now/forecasts results displayed in Table 3 are similar to the density prediction results displayed in Table 2. First, the raw RMSFE values of the BM model, which is the plain mixed frequency dynamic factor model without stochastic volatility, indicate that the data flow matters for point prediction, as in the case of density predictions. In the case of nowcasting results, under the header “Q0,” the impact of data flow is largest. This impact fades out smoothly with the increasing forecast horizon as the reduction in the RMSFEs is the smallest for the 2-quarter ahead predictions. An important finding is on the survey-based predictions. Survey participants provide quite accurate predictions with the relative RMSFEs around 86% compared to the BM models’ predictions. This performance is relatively stable and statistically significant at conventional significance levels for nowcasts and 1- and 2-quarter ahead forecasts. Hence, in line with the existing literature, SPF provides an important predictive source for point predictions based on RMSFEs.

Next, we consider the importance of allowing stochastic volatility for the point prediction. For point nowcasting, the relative RMSFEs are around 95% compared to the standard BM model without stochastic volatility. This finding shows that in the nowcasting case, the baseline statistical model of BMSV performs comparably to the SPF predictions. Hence, the base-



**Table 3.** Relative RMSFEs of competing models.

	Q0			Q1			Q2		
	M1	M2	M3	M1	M2	M3	M1	M2	M3
BM: Raw values	<b>0.705</b>	<b>0.628</b>	<b>0.610</b>	<b>0.772</b>	<b>0.787</b>	<b>0.719</b>	<b>0.787</b>	<b>0.786</b>	<b>0.777</b>
SPF		0.88		0.84			0.86		
(BM+S)-M0	0.87	0.89	0.89	0.89	0.86	0.88	0.94	0.92	0.91
(BM+S)-M1	0.87	0.89	0.89	0.89	0.85	0.87	0.93	0.91	0.91
(BM+S)-M2	0.87	0.89	0.89	0.89	0.86	0.88	0.95	0.90	0.91
(BM+S)-M3	0.87	0.91	0.92	0.89	0.86	0.88	0.95	0.91	0.91
(BM+S)-M4	0.86	0.92	0.93	0.88	0.86	0.88	0.94	0.91	0.91
BMSV	0.95	0.93	0.95	0.96	0.97	0.97	0.97	0.98	0.97
(BMSV+S)-M0	0.84	0.83	0.85	0.89	0.87	0.88	0.95	0.92	0.92
(BMSV+S)-M1	0.84	0.83	0.85	0.88	0.84	0.87	0.93	0.91	0.91
(BMSV+S)-M2	0.84	0.84	0.85	0.89	0.84	0.85	0.95	0.90	0.90
(BMSV+S)-M3	0.83	0.85	0.86	0.87	0.84	0.85	0.94	0.89	0.89
(BMSV+S)-M4	0.83	0.86	0.86	0.88	0.85	0.85	0.94	0.89	0.88
(BMSV+S)-MV0	0.84	0.82	0.84	0.88	0.86	0.87	0.93	0.92	0.92
(BMSV+S)-MV1	0.83	0.82	0.84	0.88	0.85	0.86	0.93	0.90	0.91
(BMSV+S)-MV2	0.83	0.83	0.84	0.88	0.84	0.86	0.93	0.90	0.90
(BMSV+S)-MV3	0.82	0.84	0.85	0.87	0.84	0.85	0.94	0.89	0.88
(BMSV+S)-MV4	0.83	0.84	0.86	0.87	0.84	0.85	0.94	0.90	0.89

NOTE: The Root Mean Squared Forecast Errors (RMSFE) are computed at the end of each month in the current quarter, for the current quarter (Q0-M1, Q0-M2, Q0-M3), which implies nowcasts, for one quarter ahead (Q1-M1, Q1-M2, Q1-M3) and for two quarters ahead (Q2-M1, Q2-M2, Q2-M3), which imply forecasts, over the evaluation period starting from 1977 until 2019 recursively. The estimation sample starts from 1968. BM stands for the baseline model as described in Equation (5) and (BM+S)-Mk stands for the model where we extend the BM model with the first moment of the survey-based predictions for  $k = 0, 1, \dots, 4$  as described in Equation (11). BMSV stands for the baseline model together with stochastic volatility as described in Equations (6) and (7). Finally, (BM+SV)-MVk stands for the model where we extend the BMSV model with the first and the second moment of the survey-based predictions for  $k = 0, 1, \dots, 4$  as described in Equations (12) and (14). Statistical significance of the predictive Bayes factors are tested using the Diebold–Mariano (DM) test using squared error contributions together with HAC covariance matrix and a finite sample correction, Harvey, Leybourne, and Newbold (1997). The cells with white background contain the values that are statistically *insignificant* at conventional significance level of 5%.

line statistical model already provides considerable power as a competitive benchmark. However, the performance of the BMSV model smoothly deteriorates with the increasing forecast horizon reaching 98% for the 2-quarter ahead predictions. This result implies that the importance of incorporating stochastic volatility disappears at the long horizons departing from our findings on density nowcasting in Table 2.

The point prediction capability of the (BM) BMSV models improves impressively when the survey-based expectations in terms of the first moments are incorporated into the statistical model, that is, ((BM+S)-Mk) (BMSV+S)-Mk models. In the case of nowcasting, when survey point nowcasts and forecasts are added, the relative RMSFEs are as high as 84% of the BM models' RMSFE, which are statistically significant. This result indicates that when the baseline statistical models are aligned with the survey-based predictions, the resulting model can outperform both individual sources of predictions, as can be seen in this case. We observe a similar improvement when we focus on 1-quarter ahead predictions. While the improvement, once again, is substantial compared to the BMSV model, it is pretty much similar to the SPF 1-quarter ahead predictions. For 1-quarter ahead predictions, the aligned model performs as well as the best performing predictive source, the SPF. This outperformance erodes partially with the increasing forecast horizon in the sense that for the 2-quarter ahead predictions, the relative RMSFEs are around 90%, especially when all horizons are aligned with

the baseline statistical model. Still, some of these differences are statistically significant in these cases. However, the performance of the (BMSV+S)-Mk models is worse than the SPF model predictions for the 2-quarter ahead predictions. While the (BMSV+S)-Mk models perform much better than the BMSV models, with the increasing gap between the SPF and BMSV, the ability to perform better than the SPF itself becomes harder to achieve.

Finally, we consider the (BMSV+S)-MVh models' point prediction performance. In general, for these models, we observe an increase of 1–2 points on top of the (BMSV+S)-Mk relative RMSFEs. This marginal increase shows that the alignment of the first moments gets the lion's share in improving the (BMSV+S) models performance compared to the alignment of the second moment for point predictions. On the other hand, when we consider density prediction in Table 2 we conclude that both the first and second moments are crucial for the predictive performance of the models regardless of whether we focus on the nowcasts or 1- and 2-quarter ahead forecasts.

## 5. A Closer Look at the Predictive Performance of the Models During the Covid-19 Pandemic

The outbreak of the novel coronavirus, Covid-19, has led to a dramatic health crisis. Several countries, including the United States, have taken measures to contain the pandemic. These measures include partial or complete closure of various businesses leading to a devastating supply shock, see Alvarez, Argente, and Lippi (2021); Acemoglu et al. (2021); Çakmaklı et al. (2021). On top of that, the pandemic has substantially altered daily routines with a fundamental change in preferences, leading to a sizeable demand shock, see Eichenbaum, Rebelo, and Trabandt (2021). Note that our results in previous sections are based on the dataset that excludes 2020. Here, we would like to have a closer look at the predictive ability of the competing models during the turmoil periods of 2020. In the upper panel of Table 4, we display the mean of survey participants' predictions and the (square root of) disagreement among the forecasters as computed in Equation (13) that includes the releases in the second and third quarters of 2020.<sup>17</sup> Similarly, we display the mean and the volatility of the predictive distributions obtained by the baseline statistical models, BM and BMSV, and the most general survey augment model, (BMSV+S)-MV4 in the following rows. These predictions are performed at the end of each month, as stated in the corresponding column's header. Finally, the actual realization of the real GDP growth and the point nowcasts provided by the Federal Reserve Bank of New York are displayed in the last two rows of Table 4.

First, we focus on the nowcast densities of the second-quarter GDP growth using available data at the end of April, May, and June. The economic downturn and the uncertainty brought by the devastating shock due to the Covid-19 pandemic are overwhelming. The actual growth rate for the second quarter

<sup>17</sup>Since the fourth quarter of 2020 is relatively more in parallel with the pre-pandemic periods, here we do not provide details on the fourth quarter of 2020. We provide a complete analysis, including the fourth quarter of 2020 and the corresponding nowcast distributions obtained at the end of each month in Section G of the supplementary material.

**Table 4.** The first and second moments of the predictive distributions of 2020-Q2 and 2020-Q3 real GDP growth rates based on predictions from various sources.

		2020-Q2			2020-Q3		
		April	May	June	July	August	September
SPF-May 15	Mean		−31.5		10.4		
	$\sqrt{Dis}$		7.1		11.6		
SPF-Aug 15	Mean				19.4		
	$\sqrt{Dis}$					6.1	
BM	Mean	−13.3	−26.5	−21.2	9.1	6.0	7.3
	Vol.	1.7	2.8	3.4	3.9	3.7	4.0
BMSV	Mean	−11.9	−36.5	−28.8	10.9	12.0	13.1
	Vol.	2.8	3.7	4.1	4.0	3.6	3.5
(BMSV+S)-MV4	Mean	−7.5	−32.1	−32.3	11.0	19.0	16.7
	Vol.	3.4	3.9	3.9	3.9	3.7	3.5
New York Fed		−7.8	−35.5	−16.3	16.9	15.3	14.1
Actual			−31.7			33.1	

NOTE: The first four rows display the mean of the predictions performed by the survey participants and the square root of the disagreement between those,  $\sqrt{Dis}$ , for the second and third quarters of 2020 released on May 15, 2020, and August 15, 2020, denoted as SPF-May 15 and SPF-Aug 15. BMSV stands for the baseline model with stochastic volatility as described in Equations (6) and (7). (BM+S)-MV4 stands for the model where we extend the BMSV with the first and the second moment of the survey-based predictions for  $k = 0, 1, \dots, 4$  as described in Equations (12) and (14). New York Fed stands for the Federal Reserve Banks of New York, and the numbers correspond to the point nowcasts released at the end of the month displayed on top of the table, see <https://www.newyorkfed.org/research/policy/nowcast>. Actual stands for the realization of the real GDP growth provided by the third release, respectively, see <https://www.bea.gov/news/current-releases>.

of 2020 turned out to be as sizable as  $-31.7\%$ . Considering the end of April estimates, we observe that the baseline models, BM and BMSV, anticipate the downturn to some extent in parallel with the prediction of New York Fed as  $-13.3\%$ ,  $-11.9\%$ , and  $-7.8\%$ , respectively. However, the magnitude of the contraction is still profoundly far from the actual rate. Notice that the SPF's second-quarter release is still unavailable as of April. Therefore, when predictions of the (BMSV+S)-MV4 are performed, only 1-quarter ahead predictions of the February 14 release are used for the survey information on the second quarter of 2020 real GDP growth, which is  $2.1\%$ . Consequently, the mean nowcast for the (BMSV+S)-MV4 is  $-7.5\%$ , which is close to the nowcast of the New York Fed.

The predictions of the end of May are quite decisive for almost all of the sources. First, the May 15 release of the SPF substantially updates the predictions for the second quarter of 2020 real GDP growth to  $-31.5\%$ , together with an extreme level of disagreement. In particular, participants of the SPF provide predictions ranging from  $-10\%$  at the most optimist side to  $-50.2\%$  at the most pessimist side. This sizable amount of uncertainty is also reflected in the nowcast of the New York Fed with the prediction of  $-35.5\%$  at the end of May, which is almost 20% greater (in magnitude) than the previous month's prediction. The predictive distribution provided by the BMSV model has a mean of  $-36.5\%$  (similar to New York Fed) together with a volatility level of  $3.7\%$ . Strikingly, combined with the survey information, the predictive distribution of the (BMSV+S)-MV4 model is centered around the mean of  $-32.1\%$  with a slightly higher level of volatility of  $3.9\%$  compared to the BMSV model due to the substantial disagreement in the SPF May 15 release. As a result, both distributions include the actual value inside the 95% credibility sets.

Finally, the end of June estimates exhibit another substantial shift in the location of many of the predictions. Specifically, the efforts to reopen business are reflected in the observations of various variables in May, thereby to the predictions performed using available data as of the end of June. This reversal can also be seen in the predictions of the New York Fed with substantial revisions to  $-16.3\%$ . We observe a similar revision also for the predictive distribution obtained by the BMSV model, with a distribution centered around the mean of  $-28.8\%$ . On the contrary, the (BMSV+S)-MV4 model has predictions comparable to the previous month's predictions. Thanks to the survey information, the predictive distribution is centered around the mean of  $-32.3\%$  with a volatility of  $3.9\%$ . Therefore, (BMSV+S)-MV4 leads to the most accurate prediction of this extreme downturn with a contraction of  $-31.7\%$ , by combining the survey information with the baseline statistical model.

The U.S. economy displayed a massive bounce-back in the third quarter of 2020 with a growth rate of  $33.1\%$ . This rapid upturn poses important challenges to all competing models. At the end of July, all baseline statistical models and the New York Fed nowcast a fast but relatively milder reversal in the economy with growth rates ranging from  $9.1\%$  for the BM to  $16.9\%$  for the New York Fed. As of the end of July, the third quarter release of SPF is not available. Therefore, the combined model (BMSV+S)-MV4 heavily uses the 1-quarter ahead prediction from the May 2020 release of the SPF with the predictive mean of  $10.4\%$  and substantial volatility of  $11.6\%$ . As a result, the (BMSV+S)-MV4 model predicts  $11.0\%$  with a little use of the survey. As of the end of August, the estimation results provide compelling evidence on the use of our model structure. The August release of the SPF predicts  $19.4\%$  with relatively lower disagreement than the previous quarter's release. The baseline statistical models and the New York Fed continue with predictions ranging from  $12.0\%$  ( $6.0\%$ ) for the BMSV (BM) model to  $15.3\%$  for the New York Fed. Equipped with the SPF predictions, the (BMSV+S)-MV4 model's prediction increases to  $19.0\%$ , which is closest to the actual realization together with the SPF-based expectation itself.

This demonstration represents the underlying drivers of the findings on the predictive ability of our model. On the one hand, survey-based predictions of the first moment are mostly aligned with those from the baseline statistical model. Nevertheless, it embeds relevant information, in particular for extreme cases. On the other hand, survey-based predictions of the second moment react swiftly to the changing uncertainty. This flexibility leads to a more accurate adjustment of the predictive distribution obtained using the combined model to the new conditions than the conventional statistical model.

## 6. Comparison with Alternative Methods

The findings documented in previous sections show the importance of blending survey-based information with a state-of-the-art statistical model for timely and accurate predictions of real GDP growth. We do that by combining both sources of information in a unified model ex-ante to obtain the final predictive distribution of the target variable. We provide the derivations of the implied weights of this combination for a toy model in Section F of the supplementary material. Alternatively, we could also combine these different sources of predictive information

ex-post. In this case, the baseline statistical model is first estimated to construct the predictive distributions and combined with the survey information afterward. This section provides a comparison of our methodology with these alternative ex-post combination methods of the predictive distribution obtained for the BMSV model<sup>18</sup> and the SPF information.

The methods that we consider fall into the category of forecast density combination. First, we consider (dynamic) Bayesian model averaging (BMA) using predictive likelihoods in a rolling window as a simple yet quite useful tool for density combination, see Aastveit et al. (2014). Second, we consider the forecast density combination approach with time-varying weights with proper restrictions on the weights ensuring convexity, abbreviated as DeCo, Billio et al. (2013). DeCo has the advantage of using full predictive distribution for the BMSV model obtained from the simulation scheme as observables in the forecast combination model. The final forecast combination method that we consider is the Bayesian Predictive Synthesis (BPS) approach, McAlinn and West (2019). Essentially, the foundational framework of BPS already nests the former approaches as special cases, as also noted in McAlinn and West (2019). The BPS model where we use  $h$ -quarter ahead predictions from individual models for obtaining combination density is denoted as BPS( $k$ ). While for the BPS and BMA<sup>19</sup> analytical solutions for computing the predictive likelihood are readily available, these do not exist for the remaining methods. Consequently, while we report the RMSFEs, predictive likelihoods are evaluated only for BPS, BMA, and (BMSV-S)-MV $k$  models. We display the results in Table 5.<sup>20</sup>

When we consider point forecasting results using the RMSFE metric, we observe an almost monotonic increase in the performance starting from the least sophisticated method, (dynamic) BMA to the BPS and DeCo, and finally (BMSV+S)-MV $k$  models. In case of nowcasting displayed under the column with the header Q0, the BPS model and the DeCo model perform very similarly, reflecting the similarities between the two methods. The (BMSV+S)-MV $k$  models perform better than the alternative methodologies with (BMSV+S)-MV4 model outperforming all methods with almost a 5% difference in terms of relative RMSFEs.

Table 5. Comparison of alternative methodologies.

	RMSFE			Predictive likelihoods		
	Q0	Q1	Q2	Q0	Q1	Q2
BMSV: Raw values	<b>0.57</b>	<b>0.66</b>	<b>0.63</b>	<b>−94.8</b>	<b>−113.6</b>	<b>−117.6</b>
BMA	0.998	0.991	0.998	4.61	3.21	0.33
DeCo	0.875	0.926	0.997			
BPS(0)	0.876	0.934	1.032	6.85	2.66	−10.62
BPS(k)	0.876	0.907	0.960	6.85	6.10	−1.09
(BMSV-S)-MV0	0.829	0.870	0.903	4.65	3.61	3.46
(BMSV-S)-MV1	0.830	0.848	0.933	8.05	5.45	3.64
(BMSV-S)-MV2	0.834	0.842	0.903	8.71	5.70	4.09
(BMSV-S)-MV3	0.837	0.845	0.915	8.16	5.38	4.12
(BMSV-S)-MV4	0.825	0.848	0.912	9.44	5.55	4.48

NOTE: The Root Mean Squared Forecast Errors (RMSFE) and the log-predictive likelihoods Log(-PL) are computed at the end of the second month in the current quarter (Q0), which implies nowcasts, for one quarter ahead (Q1) and for two quarters ahead (Q2), which imply forecasts, over the evaluation period starting from 1988 since these requires the estimation of an ex-post model of forecast combination. The estimation sample starts from 1968. BMSV stands for the baseline model together with stochastic volatility as described in Equations (6) and (7). Finally, (BM+SV)-MV $k$  stands for the model where we extend the BMSV model with the first and the second moment of the survey-based predictions for  $k = 0, 1, \dots, 4$  as described in Equations (12) and (14). Log-predictive likelihoods are nonexistent for DeCo due to the lack of analytic predictive distributions. The cells with white background contain the values that are statistically *Insignificant* at the conventional significance level of 5%.

Considering the competing models' density nowcasting performance, we observe a similar monotonic pattern of improvement as in the previous case. In this case, the ex-post combination of predictive nowcast densities using the BPS model improves the predictions by almost 7 points over the baseline BMSV model. The (BMSV+S)-MV $k$  models involving the ex-ante combination of predictive models further improve on top of the BPS model. In this case, the difference becomes almost 9 points over the baseline BMSV model. An important finding is on improving the density nowcasting with the inclusion of the SPF forecasts. The improvement is greatest when all available SPF nowcasts and forecasts are aligned in (BMSV+S)-MV4. This improvement shows the importance of a flexible model structure that can incorporate all available information at all horizons.

Next, we consider the findings related to 1- and 2-quarter ahead predictions. The point prediction results using the RMSFE computations reveal a very similar pattern as in the case of nowcasts. In this case, the (BMSV+S)-MV2 and the (BMSV+S)-MV0 model provides the best predictions for the 1- and 2-quarter ahead predictions measured using the RMSFE metric, respectively. An important finding is on the BPS models. While the BPS(0) model that provides 1- and 2-quarter ahead predictions with nowcasts densities using direct extrapolation performs worse, the BPS( $k$ ) model that uses individual 1-quarter ahead predictions for synthesis performs better than the DeCo model. These findings are in line with those documented in McAlinn and West (2019). Our findings are similar for the density prediction results using the (log-)predictive likelihood computations for the 2-quarter ahead forecasts but not for the 1-quarter ahead density forecasting. For the 1-quarter ahead density forecasting, the BPS( $k$ ) provides the best density prediction with a difference of almost 6 points compared to the baseline statistical model, BMSV. Still, the BPS( $k$ ) results are very close to the (BMSV+S)-MV $k$  models with differences smaller than 1 point.

<sup>18</sup>We only include the BMSV model for the sake of brevity. The evidence in earlier sections shows that the BMSV model has superior predictive ability compared to the BM model, and thus the conclusions drawn in the section remain unaffected if we also include BM in the comparison of methods.

<sup>19</sup>For BMA, we approximate the SPF predictions using a parametric specification, allowing us to compute the predictive likelihood. We provide the details on this model and on forecast combination methods in Section F of the supplementary material.

<sup>20</sup>We also consider the Entropic Tilting (ET) method in comparison to our model framework. In the ET method, the combined distribution is obtained by minimizing the relative entropy, that is, the Kullback–Leibler divergence, between the candidate distribution and the predictive distribution obtained by the BMSV model subject to the constraints that the moment conditions of the candidate should be identical to those from the SPF, see, for example, Krüger, Clark, and Ravazzolo (2017) and Tallman and Zaman (2019). Unlike our models, where the moments from both sources could occasionally deviate from each other, the moments are perfectly identical in our case for the ET methods. This equality implies that the conclusions that we draw using RMSFEs in Section 4.4 for comparison between the SPF and the BMSV model also apply here. While analytical expressions for predictive likelihoods are not available, we evaluate the ET method in comparison to our model framework using other metrics, which are displayed in Section F of the supplementary material.



## 7. Conclusion

We propose an econometric model where we incorporate the survey-based information to the conventional dynamic factor model (together with stochastic volatility) used for nowcasting the U.S. real GDP in a statistically coherent way. Our model effectively combines the predictive distributions of the real GDP implied by the predictions of survey participants with the predictive distributions implied by the statistical model. We do this by aligning the implied first and second moments of those predictive distributions from these two sources of information. The resulting approach naturally fits the state-space structure sidestepping the need for the tilting or ex-post combination methods used in similar setups.

Our model produces survey consistent measures of output growth expectations and accompanying time-varying uncertainty. We use the output projections for different horizons from the Survey of Professional Forecasters (SPF). We provide results on the accuracy of nowcasts and short-term forecasts of U.S. real GDP growth in a real-time exercise with the evaluation period from the first quarter of 1977 until the end of 2019. A comparison of different specifications through the predictive likelihoods and RMSFEs reveals the importance of the survey information in predicting the density of the U.S. real GDP. A month-by-month analysis in the turmoil periods of 2020 confirms the outperformance of the survey enhanced model in nowcasting the density of the U.S. real GDP.

## Supplementary Materials

The supplementary material contains details on the econometric model and related Bayesian inference, comparison with alternative specifications, analysis with alternative datasets, evaluation of the prediction performance using Probability Integral Transforms (PIT), comparison with alternative density combination methods, and finally, a detailed monthly analysis of the model performance during 2020.

## References

- Aastveit, K. A., Gerdrup, K. R., Jore, A. S., and Thorsrud, L. A. (2014), "Nowcasting GDP in Real Time: A Density Combination Approach," *Journal of Business & Economic Statistics*, 32, 48–68. [1,5,14]
- Abel, J., Rich, R., Song, J., and Tracy, J. (2016), "The Measurement and Behavior of Uncertainty: Evidence from the ECB Survey of Professional Forecasters," *Journal of Applied Econometrics*, 31, 533–550. [5]
- Acemoglu, D., Chernozhukov, V., Werning, I., and Whinston, M. D. (2021), "Optimal Targeted Lockdowns in a Multigroup SIR Model," *American Economic Review: Insights*, 3, 487–502. [12]
- Altavilla, C., Giacomini, R., and Ragusa, G. (2017), "Anchoring the Yield Curve Using Survey Expectations," *Journal of Applied Econometrics*, 32, 1055–1068. [2]
- Altug, S., and Çakmaklı C. (2016), "Forecasting Inflation Using Survey Expectations and Target Inflation: Evidence for Brazil and Turkey," *International Journal of Forecasting*, 32, 138–153. [2]
- Alvarez, F., Argente, D., and Lippi, F. (2021), "A Simple Planning Problem for Covid-19 Lock-Down, Testing, and Tracing," *American Economic Review: Insights*, 3, 367–82. [12]
- Ang, A., Bekaert, G., and Wei, M. (2007), "Do Macro Variables, Asset Markets, or Surveys Forecast Inflation Better?" *Journal of Monetary Economics*, 54, 1163–1212. [1]
- Aruoba, B. S. (2020), "Term Structures of Inflation Expectations and Real Interest Rates," *Journal of Business & Economic Statistics*, 38, 542–553. [2]
- Banbura, M., Giannone, D., Modugno, M., and Reichlin, L. (2013), "Now-Casting and the Real-Time Data Flow," *Handbook of Economic Forecasting*, 2, 195–237. [1,2,3,5]
- Banbura, M., and Modugno, M. (2014), "Maximum Likelihood Estimation of Factor Models on Datasets with Arbitrary Pattern of Missing Data," *Journal of Applied Econometrics*, 29, 133–160. [3]
- 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, 213–232. [2,14]
- Bloom, N. (2014), "Fluctuations in Uncertainty," *Journal of Economic Perspectives*, 28, 153–176. [1]
- Boero, G., Smith, J., and Wallis, K. F. (2008), "Uncertainty and Disagreement in Economic Prediction: The Bank of England Survey of External Forecasters," *The Economic Journal*, 118, 1107–1127. [5]
- Bok, B., Caratelli, D., Giannone, D., Sbordone, A. M., and Tambalotti, A. (2018), "Macroeconomic Nowcasting and Forecasting with Big Data," *Annual Review of Economics*, 10, 615–643. [1]
- Bomberger, W. A. (1996), "Disagreement as a Measure of Uncertainty," *Journal of Money, Credit and Banking*, 28, 381–392. [5]
- Çakmaklı, C., Demiralp, S., Kalemli-Özcan, S., Yeşiltaş, S., and Yıldırım, M. A. (2021), "The Economic Case for Global Vaccinations: An Epidemiological Model with International Production Networks," working paper, National Bureau of Economic Research, NBERw28395. [12]
- Campbell, S. D. (2007), "Macroeconomic Volatility, Predictability, and Uncertainty in the Great Moderation: Evidence from the Survey of Professional Forecasters," *Journal of Business & Economic Statistics*, 25, 191–200. [1]
- Clements, M. P. (2014), "Forecast Uncertainty—Ex ante and Ex post: US Inflation and Output Growth," *Journal of Business & Economic Statistics*, 32, 206–216. [5]
- D'agostino, A., McQuinn, K., and Whelan, K. (2012), "Are Some Forecasters Really Better than Others?" *Journal of Money, Credit and Banking*, 44, 715–732. [1]
- Doh, T., and Smith, A. L. (2018), "Reconciling VAR-based Forecasts with Survey Forecasts," KC fed research working papers 18-13, Kansas City Fed. [2]
- Eichenbaum, M. S., Rebelo, S., and Trabandt, M. (2021), "The Macroeconomics of Epidemics," *The Review of Financial Studies*, 34, 5149–5187. [12]
- Faust, J., and Wright, J. H. (2013), "Forecasting Inflation," in *Handbook of Economic Forecasting* (Vol. 2), eds. G. Elliott, and A. Timmermann, pp. 2–56, Amsterdam: Elsevier. [1]
- Forni, C., Ghysels, E., and Marcellino, M. (2013), "Mixed-Frequency Vector Autoregressive Models," *Advances in Econometrics*, 32, 247–272. [1]
- Giannone, D., Reichlin, L., and Simonelli, S. (2009), "Nowcasting Euro Area Economic Activity in Real Time: The Role of Confidence Indicators," *National Institute Economic Review*, 210, 90–97. [1]
- Giordani, P., and Söderlind, P. (2003), "Inflation Forecast Uncertainty," *European Economic Review*, 47, 1037–1059. [1]
- Grischenko, O., Mouabbi, S., and Renne, J.-P. (2019), "Measuring Inflation Anchoring and Uncertainty: A U.S. and Euro Area Comparison," *Journal of Money, Credit and Banking*, 51, 1053–1096. [1,2]
- Harvey, D., Leybourne, S., and Newbold, P. (1997), "Testing the Equality of Prediction Mean Squared Errors," *International Journal of Forecasting*, 13, 281–291. [8,9,12]
- Kozicki, S., and Tinsley, P. (2006), "Survey-Based Estimates of the Term Structure of Expected US Inflation," working paper series 46, Bank of Canada. [2]
- (2012), "Effective Use of Survey Information in Estimating the Evolution of Expected Inflation," *Journal of Money, Credit and Banking*, 44, 145–169. [2]
- Krüger, F., Clark, T. E., and Ravazzolo, F. (2017), "Using Entropic Tilting to Combine bvar Forecasts with External Nowcasts," *Journal of Business & Economic Statistics*, 35, 470–485. [2,14]
- Lahiri, K., and Sheng, X. (2010), "Measuring Forecast Uncertainty by Disagreement: The Missing Link," *Journal of Applied Econometrics*, 25, 514–538. [1]
- Marcellino, M., Porqueddu, M., and Venditti, F. (2016), "Short-Term GDP Forecasting with a Mixed-Frequency Dynamic Factor Model with Stochastic Volatility," *Journal of Business & Economic Statistics*, 34, 118–127. [1,3]



- Mariano, R. S., and Murasawa, Y. (2003), "A New Coincident Index of Business Cycles Based on Monthly and Quarterly Series," *Journal of Applied Econometrics*, 18, 427–443. [3]
- McAlinn, K., and West, M. (2019), "Dynamic Bayesian Predictive Synthesis in Time Series Forecasting," *Journal of Econometrics*, 210, 155–169. [2,14]
- McConnell, M. M., and Perez-Quiros, G. (2000), "Output Fluctuations in the United States: What has Changed Since the Early 1980's?" *American Economic Review*, 90, 1464–1476. [5]
- McCracken, M. W., and Ng, S. (2016), "Fred-md: A Monthly Database for Macroeconomic Research," *Journal of Business & Economic Statistics*, 34, 574–589. [5]
- Omori, Y., Chib, S., Shephard, N., and Nakajima, J. (2007), "Stochastic Volatility with Leverage: Fast and Efficient Likelihood Inference," *Journal of Econometrics*, 140, 425–449. [7]
- Rich, R., and Tracy, J. (2010), "The Relationship Between Expected Inflation, Disagreement, and Uncertainty: Evidence from Matched Point and Density Forecasts," *The Review of Economics and Statistics*, 92, 200–207. [5]
- (2021), "A Closer Look at the Behavior of Uncertainty and Disagreement: Micro Evidence from the Euro Area," *Journal of Money, Credit and Banking*, 53, 233–253. [5]
- Schumacher, C. (2016), "A Comparison of Midas and Bridge Equations," *International Journal of Forecasting*, 32, 257–270. [1]
- Tallman, E. W., and Zaman, S. (2019), "Combining Survey Long-Run Forecasts and Nowcasts with BVAR Forecasts Using Relative Entropy," *International Journal of Forecasting*, 36, 373–398. [2,5,14]
- Tanner, M. A., and Wong, W. H. (1987), "The Calculation of Posterior Distributions by Data Augmentation," *Journal of the American Statistical Association*, 82, 528–550. [6]