Forecasting with the help of survey information*

Most recent version

Federica Brenna† Žymantas Budrys‡

January 2025

Abstract

In this paper we propose a parsimonious way of combining survey expectations with empirical models to produce forecasts. We do so by augmenting a traditional vector autoregression model with forecasts for different variables and horizons from the ECB Survey of Professional Forecasters. The additional information improves estimation efficiency while maintaining a treatable model. In terms of forecasting performance, the gains from adding survey forecasts are greater at the one and two year ahead horizons, while they are limited at shorter horizons (below one year). Larger gains are found in terms of density performance than in terms of point. Forecasts of real GDP growth benefit the most from survey information, whereas inflation forecasts are improved the least. This latter result is partially driven by the very poor performance of SPF during the 2022 high inflation period. Forecasts for unemployment also benefit from including expectations for GDP and inflation, although not during the COVID pandemic period.

Keywords: Expectations, Forecasting, Judgement, Survey of Professional Forecasters
JEL Codes: C32, C33, C51, D84, E37

*We are grateful, for useful comments and suggestions, to Eleonora Granziera, Michele Lenza, Francesco Ravazzolo, two anonymous referees, and to seminar participants at the Bank of Lithuania and at the 44^{th} International Symposium on Forecasting. This work was partially funded by resources from the LMT (Lietuvos mokslo tarybos - Research Council of Lithuanian), funded under the post-doctoral project nr. P-PD-23-182. All remaining errors are our own.

†Lietuvos Bankas and Vilnius University

‡Lietuvos Bankas - CEFER and Vilnius University

1 Introduction

The expectations of professional forecasters are a well-studied and closely monitored variable, given their significance in inferring agents' forward-looking behaviour and thus guiding policy decisions. They can also be beneficial for informing forecasting exercises. However, there is still some debate as to how survey and model forecasts should be combined to enhance forecast performance and to what extent the inclusion of survey data improves the quality of the forecasts.

The literature has proposed various methods to this end, from entropic tilting and soft conditioning, to optimal pooling, to other ex-post model adjustments. A standard result is that, on average, surveys improve, or at least do not worsen, forecast accuracy. The amount of information from surveys incorporated into models can also affect results, as shown in Bańbura, Brenna, Paredes, and Ravazzolo (2021) and Galvão, Garratt, and Mitchell (2021): survey point forecasts help to steer the model forecasts in the right direction, whereas survey density forecasts may be too narrow and result in badly calibrated forecasts, particularly in stable times. In more turbulent times, on the other hand, respondents' "overconfidence" and judgement can discipline otherwise badly performing forecasts produced by purely backward looking models.

A common feature of most methods that incorporate survey and model information is that they do so *after* estimation. The most prominent example is tilting: given a forecast density from, say, a vector autoregression model, and given moments from a survey, the model density is reweighted so that its moments become "closer" to the desired (i.e., survey) moments. Optimal pooling is another way of combining model and survey densities post estimation, based on their predictive performance, while soft conditioning affects model forecasts by informing the path of forecasts with information from surveys.

In this paper, we explore yet another method (developed in own previous work, see Brenna & Budrys, 2024, later referred to as BB) of adding expectations to empirical models to produce forecasts, by augmenting a traditional Bayesian vector autoregression model (BVAR) with survey information. The concept of the method is to link data and survey forecasts across their term-structure as if the latter were produced using a time series model, assumed to be the one used by the respondents. In addition, we allow for deviations from the model-consistent forecasts in the form of idiosyncratic expert judgement. This assumption is based on the finding that judgment is pervasive in survey responses (ECB, 2019a, 2024; Stark, 2013) and it is also consistent with commonly used econometric techniques for generating conditional forecasts (Antolín-Díaz et al., 2021; Chan, Pettenuzzo, et al., 2024; Waggoner & Zha, 1999).

We want to understand whether modelling survey variables together with their observed counterparts can help to estimate the parameters of interest for forecasting. In particular, it is well known in the forecasting literature that uncertainty in model parameters can lead to less accurate forecasts. This is particularly relevant when examining a short time series, as is the case for the euro area, the object of our application. Combining survey forecasts with VAR-consistent forecasts, as we do in our setting, does not introduce additional parameters. For this reason, the survey information can, loosely speaking, act as additional observations and thus improve the parameter estimates and forecasts derived from the model.

To extract as much information as possible from the SPF, our model includes both fixed-horizon and fixed-event forecasts, covering horizons from one to five quarters ahead.[1] For robustness checks, we also examine longer horizons. Even with this extensive information, the number of parameters in the model does not increase with the number of forecast horizons. The inclusion of these forecasts enables more precise parameter estimation while keeping the model specification manageable. Without imposing this forecast structure, the coefficient vector would expand significantly, necessitating the use of Bayesian shrinkage, alternative model frameworks, or other computational methods for estimation (see, for example, Bańbura et al., 2010; Carriero et al., 2016, 2022). To address this, we provide an efficient Gibbs sampler that respects the model's restrictive structure and can scale well to accommodate longer forecast horizons.

In the empirical application, we use forecasts from the Survey of Professional Forecasters (SPF) for inflation and real GDP alongside the observed data for these two variables, unemployment rate and the short-term interest rate. We then look at the forecast performance of the recursively estimated BVAR for real GDP, HICP inflation and the unemployment rate, both averaged over all real-time vintages and over time. Our baseline model is a state-of-the-art BVAR á la Chan, Koop, and Yu (2024) with stochastic volatility without survey information. These models are known for their strong forecasting performance in point and density terms (Chan, 2023; Clark & Ravazzolo, 2015; Koop & Korobilis, 2013), making them a challenging benchmark to beat.

We find that adding short-term forecasts (up to one year ahead) of GDP growth and HICP inflation improves both point and density performance of the forecasts of GDP, inflation and unemployment. The larger gains are from four to eight quarters ahead horizons, particularly for the measure of relative density accuracy, the continuous ranked probability score. When looking at the cumulative performance over the evaluation sample, we find it varies through

[1] We do not, for the moment, include survey density histograms in the model. We recognise that this is an important omission, particularly for some specific periods (also in light of Bańbura et al., 2021), and reserve to incorporate it in a (near) future work.

time. Our evaluation sample runs from 2007Q1 to 2024Q1, and thus includes several crisis and hard-to-predict periods for the euro area economy. The addition of surveys does not help in forecasting the global financial crisis, but soon after, the accuracy of the survey-augmented model improves for all variables and horizons. Surveys are helpful for all three variables during the sovereign debt crisis period from 2012 to 2014, while in the COVID-19 pandemic, GDP and unemployment forecasts did not benefit from survey information, but inflation forecasts did. The last two years of the evaluation sample, marked by the high inflation period, see a better performance of the survey augmented model for GDP and unemployment, but a worse one for HICP, underlying the difficulty even for professional forecasters to predict the length and intensity of the high inflation episode.

This paper relates and contributes mainly to three strands of the literature. First, it relates to papers combining model and survey information in various ways, such as entropic tilting and soft conditioning (Galvão, Garratt, & Mitchell, 2021; Ganics & Odendahl, 2021; Krüger, Clark, & Ravazzolo, 2017; Robertson, Tallman, & Whiteman, 2005; Tallman & Zaman, 2020), optimal pooling (Bańbura, Brenna, Paredes, & Ravazzolo, 2021), and other ex-post model adjustments (Monti, 2010; Svensson, 2005). All of these papers propose post-estimation adjustment of model forecasts and do not exploit the information content of surveys to reduce estimation uncertainty, but rather their more accurate predictive performance. The paper closest to our methodology is Frey and Mokinski (2016): the authors exploit information from survey forecast by adding it in a similar way to ours when using a VAR model. We deterministically impose that the coefficients relating current and lagged data are the same as those relating forecast and current data, while the cited paper uses Bayesian methods to shrink the two types of coefficients close to each other, without imposing that they are the same ex ante. Results from this paper include the fact that the case with coefficients imposed to be equal is also the best performing in terms of forecast, giving us confidence in our approach. Frey and Mokinski (2016) incorporate only the nowcasts to the VAR, while we also add further horizons when available, increasing the information content from surveys. The empirical applications are also different: Frey and Mokinski (2016) focus on US data, while we look at the euro area SPF.

Second, our study connects to a vast literature discussing the rationality of survey expectations as well as deviations from the full information paradigm, with seminal papers including Coibion and Gorodnichenko (2012, 2015), Mankiw et al. (2003), Sims (2003), and Woodford (2013), to name a few. In contrast to these papers, we adopt a different strategy, refraining from committing to specific microfoundations underlying expectation formation processes and opting instead to model expectations in a flexible, simplified form. Given our goal of improving forecasting performance, we believe that maintaining a parsimonious model such as a VAR gives us an advantage over, say, a dynamic stochastic general equilibrium model, while recognizing the trade-off between higher accuracy due to lower parameter uncertainty and the ability to capture a richer set of dynamics.

Third, we expand the literature trying to bridge fixed horizon and fixed event forecasts. The majority of this literature focuses on the US SPF or on forecasts by Consensus Economics, notably, Hepenstrick and Blunier (2022) combine quarterly data and annual forecasts in a state-space model to reproduce the underlying path of quarterly forecasts and Knüppel and Vladu (2016) focus on an approximation based optimal weights, finding that they are different than the so-called "ad-hoc" weights most often used in the literature (both papers use the forecasts by Consensus Economics). For the Philadelphia Fed SPF, Clark, Ganics, and Mertens (2024) merge fixed horizon and fixed event forecasts in a unified framework in order to produce the underlying term structure of point and of density forecasts and Ganics, Rossi, and Sekhposyan (2020) also focus on the density fixed event forecasts and propose a density combination approach to retrieve fixed horizon density forecasts. We consider the ECB SPF forecasts, where the fixed horizon forecasts are only available for one and two years ahead, and thus a full "term structure" of the forecasts is missing. Similarly to Clark et al. (2024), we construct a term structure of forecasts by bridging the SPF forecasts available at a fixed horizon with those for a fixed event (namely, the current year, next year, or year after next) and with our BVAR model. We then obtain forecasts between horizon one and horizon twelve that are consistent with the survey forecasts, albeit derived from the model estimation.

The rest of the paper is organised as follows: Section 2 describes the model and Section 3 the data and accuracy metrics used. Section 4 presents results for the different accuracy metrics, Section 5 describes the robustness checks and Section 6 concludes.

2 Econometric framework

The model that we will augment with survey forecasts is a commonly used benchmark model in macroeconomic forecasting: a Bayesian vector autoregression (BVAR) with stochastic volatility (SV). BVAR models are widely used in central banks, given their high adaptability, interpretability, and forecasting accuracy (Angelini, Lalik, Lenza, & Paredes, 2019; Crump, Eusepi, Giannone, Qian, & Sbordone, 2021; Domit, Monti, & Sokol, 2016). The addition of stochastic volatility, which has been found to be an intrinsic characteristic of several macroeconomic time series aggregates, has also been shown to improve forecast accuracy to a large extent (see, for example, Banbura, Lenza, & Paredes, 2024; Clark & Mertens, 2023; Clark & Ravazzolo, 2015; Huber, Pfarrhofer, & Piribauer, 2020; Marcellino, Porqueddu, &

Venditti, 2016).

Our benchmark model, without any addition from SPF, is as follows:

$$y_t = c + \beta_1 y_{t-1} + ... + \beta_p y_{t-p} + A_0^{-1}\Lambda_t e_t \qquad\qquad e_t \sim \mathcal{N}(0, I_n) \tag{1}$$

where y_t is a vector of N observables, c is a vector of constants, β_k, $k = 1, ..., p$ are $N \times N$ matrices of lagged coefficients, p is the number of lags, A_0 is a matrix of impact coefficients, Λ_t collects scaling factors for the stochastic volatility of shocks and e_t is a vector of structural form disturbances. Note that we follow Chan, Koop, and Yu (2024) and allow the A_0 matrix to be full and identified, under some weak conditions discussed below. Unlike previous attempts to model stochastic volatility using the triangularisation of the error covariance matrix, this setting is invariant to the ordering of the variables.

As is standard practice in forecasting evaluation, we estimate the above model recursively across different real-time vintages and produce out-of-sample forecasts, which we then compare with those produced by the "augmented" model explained below.

2.1 A BVAR-SV model on data and forecasts

We augment the benchmark model with additional equations for each forecast horizon in the following form:

$$\begin{aligned} y_{t+h|t} &= c + \beta_1 y_{t+h-1|t} + ... + \beta_p y_{t+h-p|t} + A_0^{-1}\Lambda_{t+h|t} e_{t+h|t} \\ &\vdots \\ y_{t|t} &= c + \beta_1 y_{t-1} + ... + \beta_p y_{t-p} + A_0^{-1}\Lambda_{t|t} e_{t|t} \end{aligned} \tag{2}$$

where we use a notation corresponding to the information set available to the econometrician. In particular, some variables, such as HICP and GDP are observed with a lag, so the forecaster does not observe their current value and has to form an expectation about the nowcast, $y_{t|t}$, that may differ from the later available observation, y_t.

The representation is parsimonious and, although it does not align with any of the various micro-foundations of expectation formation outlined in the literature review, it does adhere to the iterative procedure for generating forecasts for autoregressive processes. To illustrate this, consider a process with one lag, without loss of generality:

$$y_{t+h|t} = \underbrace{\sum_{j=0}^{h} \beta^j c + \beta^{h+1} y_{t-1}}_{\text{unconditional forecast}} + \underbrace{\sum_{l=0}^{h} A_0^{-1}\beta^{h-l}\Lambda_{t+l|t} e_{t+l|t}}_{\text{judgment}} \tag{3}$$

We refer to the first part as an unconditional forecast or, in other words, the model-implied point forecast under the mean squared forecast error loss function, whereas the second element is a contribution of expert judgment, as defined in BB. For the sake of computational simplicity, we assume that the judgement is independent of the observations, y_t, and of the realised shocks, e_{t+j}:

$$\begin{aligned}&\mathbb{E}\left(e_{t+h|t} \mid y^{\{t-1:1\}}\right) = \mathbf{0}_{N\times 1}\\ &\mathbb{E}\left(e_{t+h|t}e'_{t+j} \mid y^{\{t-1:1\}}\right) = \mathbf{0}_{N}\end{aligned} \qquad \mathbb{E}\left(e_{t+h|t}e'_{t+j|t} \mid y^{\{t-1:1\}}\right) = \begin{cases}\mathbf{I}_N & \forall h = j\\ \mathbf{0}_N & \forall h \neq j\end{cases} \tag{4}$$

where h and j are non-negative. The rationale for this specification reflects two points. First, experts generate their forecasts by gathering available information and running models, but they can also enhance their predictions with a judgmental component or "technical assumptions" (Andre et al., 2022; Croushore & Stark, 2019; ECB, 2019a, 2024; Stark, 2013). Second, this approach aligns with the econometric techniques that forecasters and policymakers typically use to adjust their statistical models when making conditional forecasts in practice. The concept of constraining future shocks within a VAR framework underlies all conditional forecasting methods (Antolín-Díaz et al., 2021; Chan, Pettenuzzo, et al., 2024; Waggoner & Zha, 1999).

Another feature of our augmented model is that, in addition to observations, the SPF forecasts for different horizons are included in the estimation. The survey observations inform the parameter estimates, while the number of parameters remains unchanged. Alternatively, one could simply expand the state vector of the VAR, but this would result in a large number of parameters. For the model with 4 variables and 5 forecast horizons, the state vector has 36 elements. For the 4 lag specification, the coefficient matrix of a general model would contain 2328 parameters, which is large compared to our augmented model, which contains only 68 parameters. In BB, we use Monte Carlo simulations to confirm that the posterior standard deviation for the parameters is smaller when forecasts are added, similar to when the observed time series is longer. For this reason, augmenting the model with forecasts is, loosely speaking, adding additional observations. In both cases, less parameter uncertainty tends to translate into better forecasting performance. The focus of this study is to assess whether this holds true in practice.

The final component of the model is the specification for stochastic volatility, which is rather standard in our case:

$$\log \sigma_t \equiv \lambda_t = \rho\lambda_{t-1} + u_t \qquad\qquad u_t \sim \mathcal{N}(0, \sigma_u^2) \tag{5}$$

Log-volatilities evolve according to an independent autoregressive process of order one, with the long-run mean fixed at zero. Under this specification, and assuming that $\rho \neq 0$ and $|\rho| < 1$, Bertsche and Braun (2022) shows that the structural matrix A_0 is full and identified, subject to sign changes and column permutations.

We also define the evolution of expected volatility. For simplicity, we assume that expected volatility follows an unconditional path, i.e., $\lambda_{t+j|t} = \rho^j \lambda_{t-1}$ or $u_{t+j|t} = 0$ for all $j = 0, \dots, h$, meaning that survey respondents do not incorporate any conditionality regarding future uncertainty.

2.2 State space representation

We write and estimate the model in state-space form. This is necessary for three reasons: first, to account for the mismatch between the variable definitions of the observed SPF forecasts and those of the state vector; second, to capture missing observations; and third, to allow for measurement errors that arise in SPF responses.

$$X_t = Z_y Y_{t+h|t} + Z_o O_t \qquad\qquad o_{i,t} \sim \mathcal{N}(0, \sigma^2_{o,t}) \tag{6}$$

Equation 6 is the measurement equation that links the observed SPF projections and the observed data, X_t, to their model equivalents. Note that we allow the measurement vector to have missing observations, which are then interpolated using the simulation smoother from Durbin and Koopman (2002). Missing observations arise because some SPF responses are for fixed event forecasts. For example, for each quarter in 2023, the current year's fixed-event forecasts refer to the year-end value. Based on our model, this can be captured so that in each quarter there is one measurement with a different associated horizon, in 2023Q1 we have a measurement $x_{t+3|t}$, in Q2 $x_{t+2|t}$, in Q3 $x_{t+1|t}$ and, lastly, in Q4 $x_{t|t}$. In addition, to rigorously account for the information available to the forecaster, we allow for missing observations at the end of the sample due to the irregular availability of variables, also known as the "ragged edge".

Z_y is a constant and known sparse mapping matrix that we discuss in section 3.3. O_t is a vector of measurement errors. Note that we allow only some observations to be measured with error, therefore we introduce another sparse matrix, Z_o. Measurement errors, $o_{i,t}$ are independent across measurements, normally distributed but with heteroskedasticity. In this respect, we follow Clark et al. (2024), who provide a compelling and comprehensive discussion of the reasons for introducing measurement error for fixed-event SPF responses. In short, possible discrepancies in reported forecasts due to averaging, rounding or calculation can lead to volatile imputations with spillover effects on parameter estimates, latent states and hence forecasts. As suggested by the authors, we also use a horseshoe specification for measurement errors, which inherently assumes that deviations from the measurement are small, but can occasionally be very large.

$$
\begin{aligned}
Y_{t+h|t} &= C + BY_{t+h-1|t-1} + F\left(I_{H+1} \otimes A_0^{-1}\right)\Lambda_{t+h|t}\xi_{t+h|t} \\
\xi_{t+h|t} &\sim \mathcal{N}(0_{N(H+1)\times 1}, I_{N(H+1)})
\end{aligned}
\tag{7}
$$

Equation 7 presents the state transition. Intuitively, the equation collects expressions of equation 3 across survey responses and across horizons as well as observed data, equation 1, and presents them in a companion form. As above, we define N as the number of variables, H as the number of forecast horizons and p as the chosen lag length. In the system, there are $N(H+1)$ structural shocks that explain the forecast structure, collected in the vector $\xi_{t+h|t}$. The state vector, $Y_{t+h|t}$, is of size $N(H+p)\times 1$ and collects conditional forecasts, nowcasts, data and lags. The matrices B and F capture the dynamic responses. The vector of constants C collects the deterministic component over the forecast horizons. Finally, the expected and current scaling factors for the stochastic volatility of structural shocks are collected in $\Lambda_{t+h|t}$.

2.3 Bayesian estimation and priors

We estimate the main model with four variables at quarterly frequency, with GDP and HICP in log-differences, and unemployment and short-term interest rate in levels. Our baseline specification includes four lags.

We use Bayesian inference with fairly standard Minnesota priors, briefly described below.[2] The model is estimated using Markov Chain Monte Carlo (MCMC) algorithm of Gibbs sampling. Due to limited data availability, we estimate the model recursively, adding one quarter at each estimation, starting from 2007Q1, up to 2024Q3, for a total of 71 vintages. At each estimation, we produce $3\cdot 10^5$ draws from the algorithm, of which $2\cdot 10^5$ are for burn-in, and we retain one in ten draws from the remaining ones, for a total of 10^4 draws for each estimation and for each forecast horizon, from horizon 0 (the nowcast) to horizon 8 (two years ahead).

Our priors are mostly uninformative but proper. For the coefficient vector, $vec(\beta)$, we set a rather uninformative Minnesota prior in the specification with average SPF forecasts (Litterman, 1986). Priors for the parameters that govern the law of motion of the stochastic volatility, ρ and σ_u^2, are independent across different processes and conditionally conjugate. For the latter parameter we use a non-standard hierarchical setup that allows for a 'fatter'

[2]For a more detailed description of the priors and estimation procedure, the reader may refer to our previous paper, Brenna and Budrys (2024), particularly its Appendix B.

tail for values close to zero and which does not rule out homoskedasticity a priori.[3] The prior for the initial state of the log standard deviation, λ_1, is hierarchical and set to the long-run values of an AR(1) process with the mean value of zero and variance determined by hyper-parameters ρ_i and $\sigma^2_{u,i}$. Priors for the initial conditions Y_0 are centred around observed data prior to the estimation period; the variance matrix is a scaled matrix, $\Sigma_{Y,0}$ which has on the diagonal the long-run variances of independent AR(4) processes over the entire sample.

3 Data

We use two main sources of data in the analysis: real time vintages from the Real Time Database by Giannone et al. (2012) and the ECB-SPF.

3.1 Real time data

To maintain the out-of-sample nature of our analysis, we collect real time vintages from the ECB Statistical Data Warehouse for four variables: unemployment rate (as a % of labour force), real GDP, HICP and 3-month Euribor. We seasonally adjust HICP using the TRAMO-X13 package developed by Lengwiler (2017), based on the method by Monsell (2007). We then convert the three monthly variables (unemployment, HICP and Euribor, the latter converted from daily frequency) to quarterly frequency, using the average of the quarter for HICP and unemployment, and the last value of the quarter for the 3-month Euribor. Table 1 summarizes the series used and transformations applied, as well as the available sample.

Table 1: Dataset for the baseline model

Variable	Transformation	Aggregation	Identifier	Start	Mnemonic
Unemployment rate	level	Average of Q	RTD.M.S0.S.L_UNETO.F	2000Q1	UNEMP
Real GDP	Δlog-level	-	RTD.Q.S0.S.G_GDPM_TO_C.E	1995Q1	GDP
HICP	Δlog-level, SA	Average of Q	RTD.M.S0.N.C_EUR3M.E	1996Q1	HICP
Euribor 3-month	level	End of Q	RTD.M.S0.N.P_C_OV.X	1994Q1	STN

Note: All series taken from the ECB Real Time Database. Real GDP already at quarterly frequency.

The purpose of our analysis is to quantify the added value of information from the SPF for a model forecast. Accordingly, we select data vintages available just after SPF cut-off dates, that is, the dates by when respondents need to reply to the survey. We can reasonably assume that respondents will have used all information available by that date, and that

[3] Our specification provides a computationally convenient prior adhering to remarks by Gelman (2006) and Kastner and Frühwirth-Schnatter (2014) that a standard prior of inverse gamma tends to be over informative a posteriori.

information from the respective SPF round will become available to policy-makers such as ECB's and other national banks' economists involved in the forecast production. That way, we can "simulate" an environment where policy-makers producing forecasts have the choice to include the latest SPF in their model, and we can verify whether doing so would improve accuracy with respect to the baseline BVAR-SV model.

Given the limited length of the sample, we estimate the model recursively, at first from 2000Q1 to 2007Q1 and then adding one quarter at each subsequent vintage. We estimate and build forecasts for a total of 71 vintages, though due to data availability for the last few vintages, we can only evaluate the forecast performance of up to $71 - H$ vintages (i.e., 70 vintages for the one quarter ahead forecasts, 63 vintages for the eight quarter ahead forecasts). The forecasts are evaluated with respect to the second release of the corresponding variable (e.g., if we are forecasting GDP for 2020Q1, we assess our forecasts against the vintage published in 2020Q3).

3.2 ECB-SPF

The second source of information is the Survey of Professional Forecasters, published quarterly by the European Central Bank since 1999 (ECB, 2019b). The purpose of the survey is to assess the expectations of experts regarding inflation and other macroeconomic variables such as GDP and unemployment for several future horizons. Respondents are members from financial and non-financial institutions, as well as research centres and institutes. Individual responses are anonymised and publicly available, and an average of 55 experts fill in the survey each quarter. In this analysis we are focusing on the average responses, calculated as the arithmetic mean of all individual responses at each quarter and for each forecast horizon.[4]

Since we want to use all available information, we decide to include in the model both "fixed horizon" forecasts (for one and two years ahead) and "fixed event" forecasts (for next year and the year after next). We do not include longer-term forecasts, such as four and five years ahead, which obviously vary over time. This is at odds with the model, that assumes a stable long-term average.[5] Merging fixed event and fixed horizon forecasts and incorporating them into our model in a consistent way is a challenge in itself, so we devote the next subsection and part of the appendix to it.

[4] While renouncing for the time being to the interesting features of heterogeneity across agents, we decide to use the most widely reported and "hard-to-beat" quantity (Genre et al., 2013), namely the average, and plan to focus on individual responses in future work.

[5] Allowing for the time-varying long-term mean to fit long-term expectations is an interesting avenue for future research. Bańbura and van Vlodrop (2018) provides compelling evidence of the good forecasting performance of a model that allows for long-term survey forecasts to discipline the time-varying mean.

3.3 Merging information from all horizons

Our goal is to convert (annual) fixed event forecasts into (quarterly) fixed horizon forecasts so that they are comparable and can be easily incorporated into our model, together with the already available fixed horizon forecasts. Since annual fixed event forecasts are repeated at each quarter with the same reference year, in the case of year-on-year growth rates it is possible to convert them to quarterly frequency by using consecutive quarterly growth rates:

$$x_t = \frac{1}{4}\left(y_{t,4} + 2y_{t,3} + 3y_{t,2} + 4y_{t,1} + 3y_{t-1,4} + 2y_{t-1,3} + y_{t-1,2}\right) + o_t \tag{8}$$

where x_t is the annual growth rate for year t, and each $y_{t,q}$ term on the right hand side is a quarterly growth rate, where the first subscript refers to the year and the second one to the quarter to which the rate refers to (e.g., $y_{t,4}$ is the growth rate for the fourth quarter of year t, calculated as a growth rate of the level in quarter 4 with respect to the level in quarter 3 of year t); o_t is an approximation error.

Several papers check the performance of this standard approximation or propose alternative ways, (see, for example, Clark et al., 2024; Hepenstrick & Blunier, 2022; Knüppel & Vladu, 2016). Here we approximate quarterly forecasts from annual ones based on each variable's definition, as is provided in the ECB SPF documentation. For example, annual real GDP is defined as the growth rate of the average level of GDP in year t with respect to the average level in year $t-1$. Annual HICP inflation is defined similarly to GDP, while annual unemployment is given by the annual average level for the year in question. Unlike the Philadelphia Fed SPF, we do not have consecutive quarterly forecasts for the first five forecast horizons, therefore we have no way of checking consistency between observed quarterly forecasts and quarterly forecasts approximated by annual ones. Similarly to Clark et al. (2024), we assume quarterly fixed horizon forecasts are observed without measurement error, whereas annual fixed event forecasts converted to quarterly frequency suffer from an approximation error. The majority of the error comes from an approximation as the one used in Aruoba (2020) to go from arithmetic mean to geometric mean, and from approximating growth rates with natural logarithms.

It is interesting to note that the literature we consider here uses only the average response across forecasters. This fact may give rise to an additional source of inconsistency between (average) quarterly forecasts and (average) annual forecasts. If an individual respondent only provides an answer for the fixed event (or only for the fixed horizon), the average response will incorporate this difference, possibly overestimating the inconsistency between quarterly and annual forecasts.

Appendix B details the variables' definitions and transformations used to incorporate fixed event forecasts into the model.

4 Forecast evaluation for different accuracy metrics

As mentioned above, we estimate both the baseline BVAR and the survey-augmented BVAR (S-BVAR) recursively, with the first vintage going from 2000Q1 to 2006Q4, the second from 2000Q1 to 2007Q1, and so on, until 2024Q2. We evaluate forecasts from the nowcast horizon ($h = 0$) to the two years ahead horizon ($h = 8$), up to when the target is available. In the following, we present results for our three target variables, namely real GDP growth, HICP inflation and unemployment. We look at average relative point and density accuracy, at cumulative point and density accuracy over the sample, and at calibration. In the next subsection, we describe more in detail the measures used.

4.1 Accuracy metrics

We compare the models' performance across several dimensions, listed below with their respective descriptions.

Root mean square forecast error (RMSFE)

The relative RMSFE checks average point accuracy over the sample:

$$RMSFE_h = \frac{RMSFE_h^1}{RMSFE_h^2}$$

With $RMSFE_h^i$, for $i = 1, ..., M$:

$$RMSFE_h^i = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(y_t - \hat{y}_{t|t-h}^i)^2}\,,$$

where T denotes the size of the evaluation sample, and $\hat{y}_{t|t-h}^i$ denotes the median of the predictive density for y_t given the data up to $t - h$ for model i. The lower the RMSFE, the better the point accuracy. In relative terms, a relative RMSFE below 1 would indicate a more accurate model at the numerator with respect to the one at the denominator.

Continuous ranked probability score (CRPS)

The relative CRPS checks average density accuracy over the sample:

$$CRPS_h = \frac{CRPS_h^1}{CRPS_h^2}$$

With $CRPS_h^i$, for $i = 1, ..., M$:

$$CRPS_h^i = \frac{1}{T}\sum_{t=1}^{T}\left(\int_{-\infty}^{\infty}(F(y;Y_{t-h},\ldots,Y_1,M_i) - \mathbb{I}(y_t \leq y))^2 dy\right),$$

where $F(\cdot;Y_{t-h},\ldots,Y_1,M_i)$ denotes the predictive cumulative distribution function (corresponding to the predictive density $p(\cdot;Y_{t-h},\ldots,Y_1,M_i)$) and $\mathbb{I}(\cdot)$ is an indicator function. The CRPS gives a measure of the distance of the cumulative density of the forecast from the realisation; the lower the score, the better the model accuracy. A relative CRPS below 1 indicates that the model at the numerator is more accurate than the one at the denominator.

Probability integral transforms (PITs)

The PIT assesses absolute performance in terms of the model calibration:

$$PIT_t = F(y_t;Y_{t-h},\ldots,Y_1,M_i), \qquad t = T_1,\ldots,T_2\,,$$

if the predictive distribution is a good approximation of the actual distribution, the sequence $PIT_{T_1},\ldots,PIT_{T_2}$ is uniformly distributed over the interval $[0,1]$. We check the hypothesis of uniformity by performing the test of Berkowitz (2001).

Forecast performance over the evaluation sample

In addition to the metrics over the average sample, we look at relative cumulative scores, allowing us to further examine the presence of time variation in the performance of the model augmented with SPF.

Results for all metrics and selected horizon are presented in the next subsections.

4.2 Average accuracy over the sample

Results for the average accuracy over the sample are presented in Table 2. Variables and scores are on the rows and forecast horizons are on the columns. The scores are calculated by dividing the augmented, S-BVAR, scores by the traditional BVAR scores. For both RMSFE and CRPS, a number smaller than one indicates a better performance of the S-BVAR over the BVAR. A first result that stands out is that including SPF in the model does not help at very short horizons: the accuracy of the nowcasts and of the one quarter ahead forecasts are worse for all variables with the exception of GDP. A second point to note are the large gains in accuracy for unemployment: from two quarters ahead, all the way to eight quarters ahead, gains are around 6-10%. This is particularly interesting considering that we are incorporating survey information only about GDP and inflation. For the latter two variables, results are mixed: GDP forecasts improve marginally in the point dimension, but to a greater extent in the density dimension. For all forecast horizons, the S-BVAR provides a lower CRPS than the BVAR model, with gains up to 11%. HICP inflation seems to be the variable benefitting the least from the addition of survey information, but it is nevertheless worth mentioning that both point and density metrics indicate an improvement in forecast accuracy for the higher horizons, between 5 and 8 quarters ahead.

Table 2: Relative accuracy scores

Horizon	h=0	h=1	h=2	h=3	h=4	h=5	h=6	h=7	h=8
GDP									
RMSFE	1.05	**0.99**	1.01	**0.99**	1.00	1.00	**0.99**	**0.99**	**0.98**
CRPS	**0.98**	**0.95**	**0.94**	**0.90**	**0.94**	**0.94**	**0.91**	**0.89**	**0.92**
HICP									
RMSFE	1.23	1.05	1.03	**0.96**	1.01	**0.98**	**0.98**	**0.97**	**0.98**
CRPS	1.13	1.03	1.00	**0.96**	1.01	**0.97**	**0.95**	**0.94**	**0.95**
Unemp.									
RMSFE	1.07	1.03	**0.96**	**0.91**	**0.90**	**0.90**	**0.90**	**0.90**	**0.90**
CRPS	1.02	1.04	**0.98**	**0.94**	**0.93**	**0.93**	**0.93**	**0.93**	**0.93**

Note: The table shows accuracy metrics in relative terms, with the BVAR model at the denominator and the Survey-augmented BVAR (S-BVAR) at the numerator. Values smaller than one indicate that the latter model is more accurate on average over the sample for the respective horizon.

4.3 Accuracy over time

The average scores could mask a lot of variation over the sample, which as mentioned includes periods of very different degrees of predictability (from the more stable low inflation period, characterising the euro area economy from 2014 till before the COVID pandemic, all the way to the unexpected high inflation episode of 2021 and 2022). For this reason, in Figure 1 we present cumulative relative scores for two selected horizons, four and eight quarters ahead.[6] The scores are calculated as the cumulative sum of the difference between the two models' scores, with the BVAR first and the S-BVAR second. A decrease in the cumulative scores indicates an improvement in accuracy of the S-BVAR, and vice versa.

A clear pattern emerges from the different sub-panels: the inclusion of surveys has consistently improved forecasts accuracy over the sample, as the downward trending lines show. There are nevertheless some exceptions and times where the SPF was worsening performance. For inflation, this is very evident for both metrics and horizons shown in the figure: performance of the S-BVAR deteriorated strongly since the 2021-2022 forecasts, with a slight improvement in the last year for the four quarters ahead. Real GDP and unemployment forecast performances follow a similar pattern, with the worst performance in the period of the great financial crisis, and a slight worsening during the COVID pandemic. Interestingly,

[6]Plots for all the horizons are available in Appendix E.

the accuracy of unemployment forecasts did not deteriorate in the very end of the sample, but it has kept improving after the COVID pandemic. This is particularly remarkable given that our S-BVAR model is augmented only with SPF forecasts of GDP and inflation, which seem to be beneficial to unemployment forecasts.

4.4 Calibration

Calibration is another important aspect of density forecasts: ideally, the policymaker wants to produce forecasts whose densities are consistent with the "true" density, so as to have an accurate measure of the uncertainty surrounding the median forecast. We look at the PITs, which should ideally be uniformly distributed when the density is well calibrated, at the Berkowitz test (Berkowitz, 2001) of good calibration and at the (Rossi & Sekhposyan, 2019) test of correct specification. The Berkowitz test rejects the null hypothesis of good calibration for both the BVAR and the S-BVAR densities at most horizons. The Rossi and Shekposyan test, on the other hand, rejects the null hypothesis of correct specification at all horizons for GDP and at 7 and 8 quarters ahead for HICP and unemployment for the BVAR without survey forecasts. For the S-BVAR, the null hypothesis cannot be rejected for GDP and HICP at shorter horizons, and for unemployment at all horizons, therefore the survey-augmented model is well calibrated in more cases than the baseline BVAR, see Figures 2 and 3.

5 Robustness checks

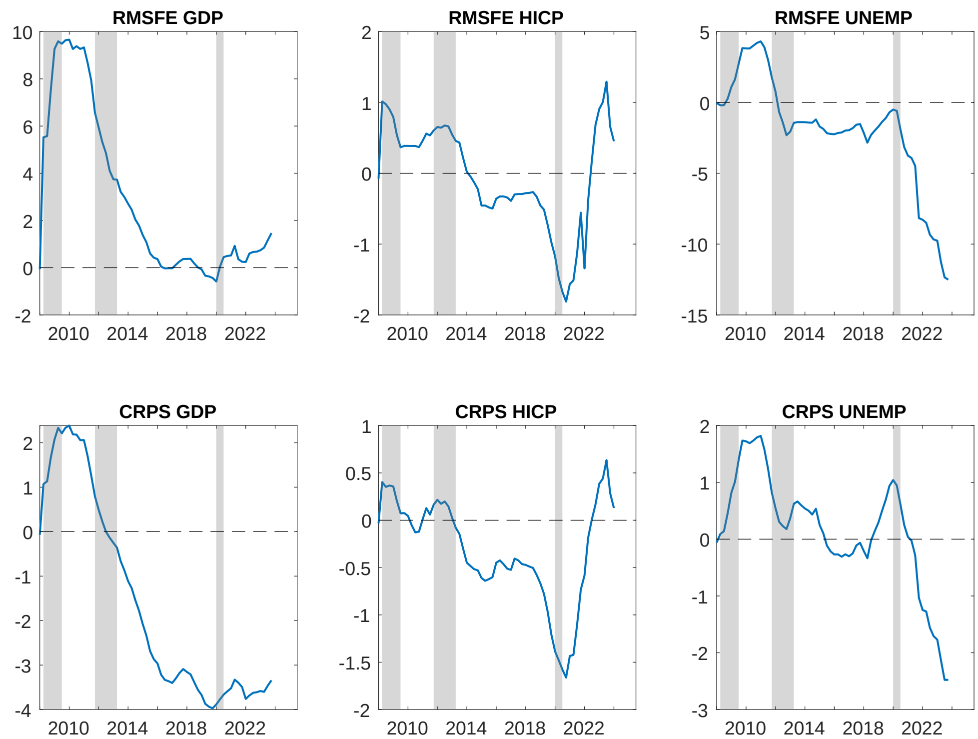
In light of the fact that our baseline exercise is a comparison between two distinct models, namely a "simple" BVAR and its augmented version incorporating survey forecasts, we conduct a series of robustness checks to ensure that our findings are not overly influenced by the selected specification. The main checks and their results are listed below.

We include four lags in our main specification, as is customary for models at quarterly frequency, but we also check results for the two lags model, suggested by one of the selection criteria. Results from the two lags models are broadly in line with the baseline, however the difference in accuracy between BVAR and S-BVAR is decreased.

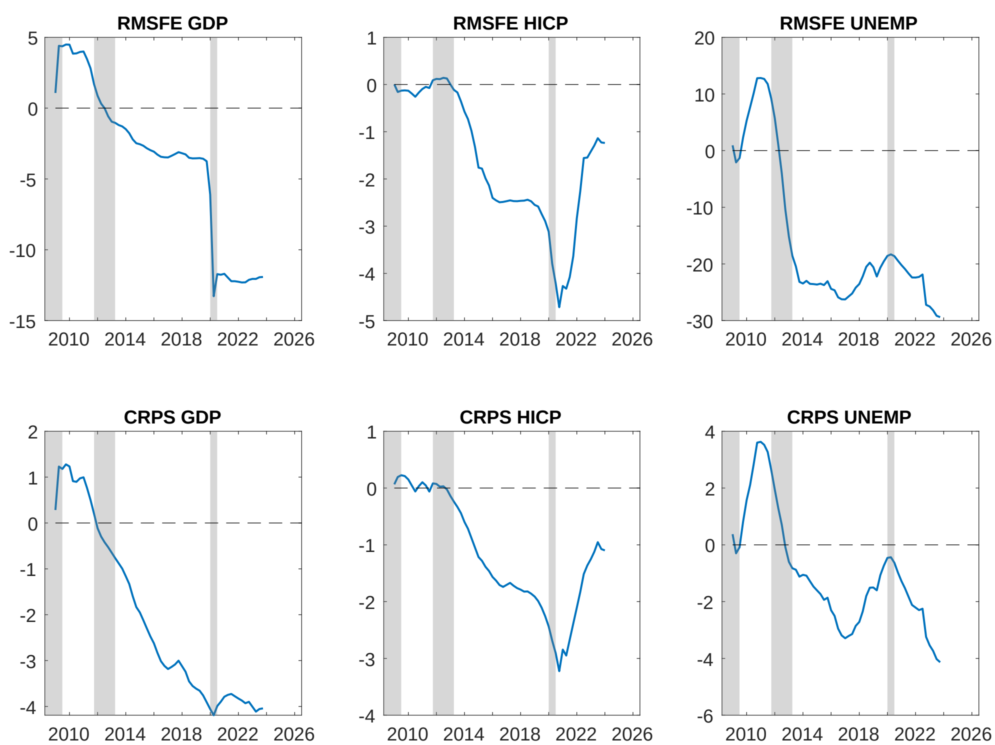
We try different priors for our parameters estimation besides the Minnesota of our main specification, namely we run the models with an adaptive hierarchical prior á la Chan, 2021. Results are comparable to the standard Minnesota case.

Additionally, we investigate the impact of incorporating varying levels of information from the SPF. Our findings suggest that including unemployment forecasts may not be beneficial

Figure 1: Cumulative relative scores for GDP, HICP and unemployment

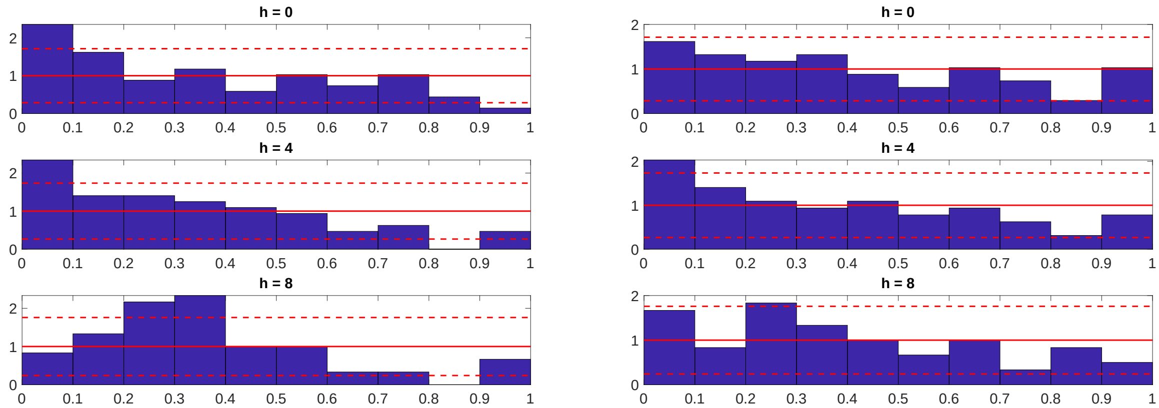


(a) Four-quarter-ahead



(b) Eight-quarter-ahead

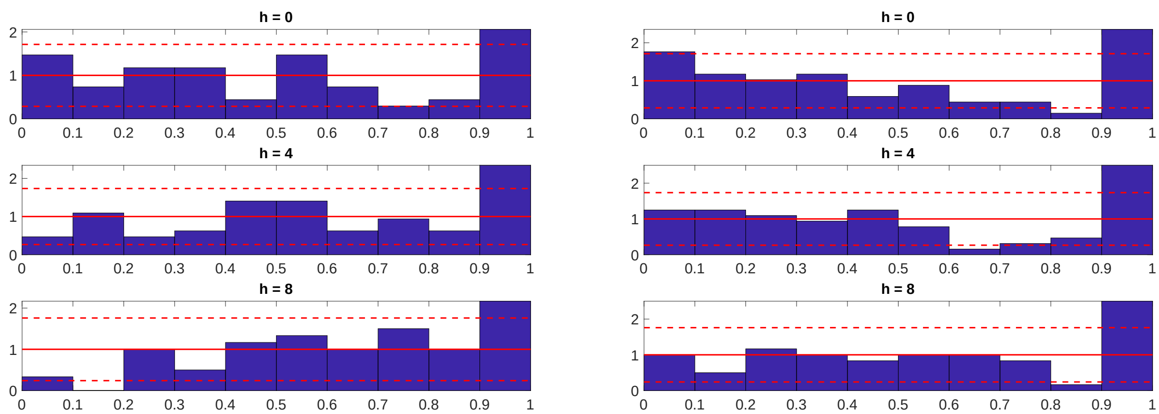
Figure 2: Probability Integral Transforms for Real GDP growth



(a) BVAR

(b) S-BVAR

Figure 3: Probability Integral Transforms for unemployment rate

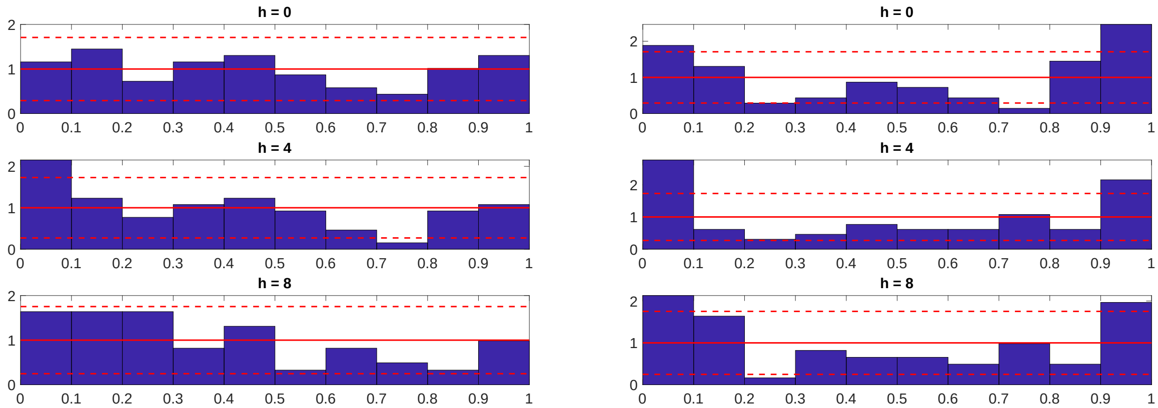


(a) BVAR

(b) S-BVAR

for overall forecast accuracy. Using forecasts of only one variable, be it GDP, inflation, or unemployment, is not enough to "beat" the baseline BVAR model. Other variables combinations, such as forecasts of GDP and unemployment or of inflation and unemployment, only do better at very short (nowcast) horizons. Including longer-term horizons from the SPF does not seem to enhance the overall performance of the forecasts. Finally, we look at the use of only the fixed horizon (one and two years ahead) or only the fixed event forecasts, in order to understand if including forecasts at different horizons does indeed improve accuracy. We find that fixed horizon forecasts alone are not sufficient to achieve the accuracy gains from the baseline case, and the inclusion of only fixed event forecasts also worsens the baseline results, suggesting that our approach of including both types of forecasts is indeed an improvement.

Figure 4: Probability Integral Transforms for HICP inflation



(a) BVAR (b) S-BVAR

6 Conclusions

In this paper we propose a new way of incorporating survey information into time series models by using survey forecasts directly in the estimation of the model. This method is a parsimonious and simple way to exploit the wealth of information contained in surveys such as the ECB-SPF. In line with the literature, we find that too much information from surveys is not always useful. We obtain improvements in forecast accuracy when adding to a four-variable VAR with stochastic volatility ECB-SPF average forecasts for HICP and GDP growth, up to the one year ahead horizon. Survey forecasts about higher horizons and about unemployment, if added to the model, are worsening overall accuracy. Interestingly, the improvement in overall forecast performance in the model augmented with selected survey information is larger for GDP growth and for unemployment, at horizons above one year. Improving forecasts' accuracy is a constant challenge, and selecting the right information to do so is not a trivial task.

The flexibility of our model opens up several avenues for future research. We are working on evaluating the forecasting performance of the model in other settings. In particular, it is an open question whether survey information can be a useful addition in a large VAR setting with more predictors. The panel dimension of surveys provides another wealth of information that can also improve parameter estimates and hence forecasts. The ECB-SPF also provides forecasts about the so-called assumptions used by the respondents to produce their forecast, another plausible source of additional insight in the forecasters' process. Finally, information on respondent uncertainty contained in SPF histograms is another potential way of disciplining model forecasts, if incorporated with due care and to the right extent.

References

Andre, P., Pizzinelli, C., Roth, C., & Wohlfart, J. (2022). Subjective models of the macroeconomy: Evidence from experts and representative samples. *The Review of Economic Studies*, rdac008.

Angelini, E., Lalik, M., Lenza, M., & Paredes, J. (2019). Mind the gap: A multi-country BVAR benchmark for the Eurosystem projections. *International Journal of Forecasting*, *35*(4).

Antolín-Díaz, J., Petrella, I., & Rubio-Ramírez, J. F. (2021). Structural scenario analysis with SVARs. *Journal of Monetary Economics*, *117*, 798–815.

Aruoba, B. S. (2020). Term structures of inflation expectations and real interest rates. *Journal of Business & Economic Statistics*, *38*(3), 542–553.

Banbura, M., Lenza, M., & Paredes, J. (2024). 9. Forecasting inflation in the US and in the euro area. *Handbook of Research Methods and Applications in Macroeconomic Forecasting*, 218.

Bańbura, M., Brenna, F., Paredes, J., & Ravazzolo, F. (2021). *Combining bayesian vars with survey density forecasts: Does it pay off?* (Working Paper Series No. No. 2543). European Central Bank.

Bańbura, M., Giannone, D., & Reichlin, L. (2010). Large bayesian vector auto regressions. *Journal of Applied Econometrics*, *25*(1), 71–92.

Bańbura, M., & van Vlodrop, A. (2018). Forecasting with bayesian vector autoregressions with time variation in the mean. *SSRN Electronic Journal*.

Berkowitz, J. (2001). Testing density forecasts, with applications to risk management. *Journal of Business & Economic Statistics*, *19*(4), 465–474.

Bertsche, D., & Braun, R. (2022). Identification of structural vector autoregressions by stochastic volatility. *Journal of Business & Economic Statistics*, *40*(1), 328–341.

Brenna, F., & Budrys, Z. (2024). *The term structure of judgement: Interpreting survey disagreement* (Working Paper Series No. No. 123). Bank of Lithuania.

Carriero, A., Chan, J., Clark, T. E., & Marcellino, M. (2022). Corrigendum to "large bayesian vector autoregressions with stochastic volatility and non-conjugate priors" [j. econometrics 212 (1) (2019) 137–154]. *Journal of Econometrics*, *227*(2), 506–512.

Carriero, A., Clark, T. E., & Marcellino, M. (2016). Common drifting volatility in large bayesian VARs. *Journal of Business & Economic Statistics*, *34*(3), 375–390.

Chan, J. (2021). Minnesota-type adaptive hierarchical priors for large bayesian VARs. *International Journal of Forecasting*, *37*(3), 1212–1226.

Chan, J. (2023). BVARs and stochastic volatility.

Chan, J., Koop, G., & Yu, X. (2024). Large order-invariant Bayesian VARs with stochastic volatility. *Journal of Business & Economic Statistics*, *42*(2), 825–837.

Chan, J., Pettenuzzo, D., Zhu, D., & Poon, A. (2024). Conditional forecasts in large bayesian vars with multiple equality and inequality constraints.

Clark, T. E., Ganics, G., & Mertens, E. (2024). *Constructing fan charts from the ragged edge of SPF forecasts* (Working Paper Series No. No. 22-36R). Federal Reserve Bank of Cleveland.

Clark, T. E., & Mertens, E. (2023). Stochastic volatility in bayesian vector autoregressions. In *Oxford research encyclopedia of economics and finance*. Oxford University Press.

Clark, T. E., & Ravazzolo, F. (2015). Macroeconomic forecasting performance under alternative specifications of time-varying volatility: Forecasting performance and specifications of time-varying volatility. *Journal of Applied Econometrics*, *30*(4), 551–575.

Coibion, O., & Gorodnichenko, Y. (2012). What can survey forecasts tell us about information rigidities? *Journal of Political Economy*, *120*(1), 116–159.

Coibion, O., & Gorodnichenko, Y. (2015). Information rigidity and the expectations formation process: A simple framework and new facts. *American Economic Review*, *105*(8), 2644–2678.

Croushore, D., & Stark, T. (2019). Fifty years of the survey of professional forecasters. *Economic Insights*, *4*(4), 1–11.

Crump, R. K., Eusepi, S., Giannone, D., Qian, E., & Sbordone, A. M. (2021). A large bayesian VAR of the united states economy. *SSRN Electronic Journal*.

Domit, S., Monti, F., & Sokol, A. (2016, January). *A Bayesian VAR benchmark for COMPASS* (Bank of England working papers No. 583). Bank of England.

Durbin, J., & Koopman, S. J. (2002). A simple and efficient simulation smoother for state space time series analysis. *Biometrika*, *89*(3), 603–615.

ECB. (2019a). The ECB survey of professional forecasters. Forecast processes and methodologies: Results of the 2018 special survey. *The ECB Survey of Professional Forecasters (SPF)*.

ECB. (2019b). Twenty years of the ECB survey of professional forecasters. *European Central Bank*.

ECB. (2024). The ECB survey of professional forecasters. Forecast processes and methodologies: Results of the 2023 special survey. *The ECB Survey of Professional Forecasters (SPF)*.

Frey, C., & Mokinski, F. (2016). Forecasting with bayesian vector autoregressions estimated using professional forecasts. *Journal of Applied Econometrics*, *31*(6), 1083–1099.

Galvão, A. B., Garratt, A., & Mitchell, J. (2021). Does judgment improve macroeconomic density forecasts? *International Journal of Forecasting*, *37*(3), 1247–1260.

Ganics, G., & Odendahl, F. (2021). Bayesian VAR forecasts, survey information, and structural change in the euro area. *International Journal of Forecasting*, *37*(2), 971–999.

Ganics, G., Rossi, B., & Sekhposyan, T. (2020). From fixed-event to fixed-horizon density forecasts: Obtaining measures of multihorizon uncertainty from survey density forecasts. *Journal of Money, Credit and Banking*.

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models (comment on article by browne and draper). *Bayesian Analysis*, *1*(3), 515–534.

Genre, V., Kenny, G., Meyler, A., & Timmermann, A. (2013). Combining expert forecasts: Can anything beat the simple average? *International Journal of Forecasting*, *29*(1), 108–121.

Giannone, D., Henry, J., Lalik, M., & Modugno, M. (2012). An area-wide real-time database for the euro area. *Review of Economics and Statistics*, *94*(4), 1000–1013.

Hepenstrick, C., & Blunier, J. (2022). *What were they thinking? estimating the quarterly forecasts underlying annual growth projections* (Working Paper Series No. No. 5). Swiss National Bank.

Huber, F., Pfarrhofer, M., & Piribauer, P. (2020). A multi-country dynamic factor model with stochastic volatility for euro area business cycle analysis. *Journal of Forecasting*, *39*(6), 911–926.

Kastner, G., & Frühwirth-Schnatter, S. (2014). Ancillarity-sufficiency interweaving strategy (ASIS) for boosting MCMC estimation of stochastic volatility models. *Computational Statistics & Data Analysis*, *76*, 408–423.

Knüppel, M., & Vladu, A. (2016). *Approximating fixed-horizon forecasts using fixed-event forecasts* (Discussion Paper Series No. No. 28). Bundesbank.

Koop, G., & Korobilis, D. (2013). Large time-varying parameter VARs. *Journal of Econometrics*, *177*(2), 185–198.

Krüger, F., Clark, T. E., & Ravazzolo, F. (2017). Using entropic tilting to combine BVAR forecasts with external nowcasts. *Journal of Business & Economic Statistics*, *35*(3), 470–485.

Lengwiler, Y. (2017). X-13 Toolbox for Matlab, Version 1.32. *Mathworks File*.

Litterman, R. B. (1986). Forecasting with bayesian vector autoregressions: Five years of experience. *Journal of Business & Economic Statistics*, *4*(1), 25–38.

Mankiw, N. G., Reis, R., & Wolfers, J. (2003). Disagreement about inflation expectations. *NBER Macroeconomics Annual*, *18*, 209–248.

Marcellino, M., Porqueddu, M., & Venditti, F. (2016). Short-term gdp forecasting with a mixed-frequency dynamic factor model with stochastic volatility. *Journal of Business & Economic Statistics*, *34*(1), 118–127.

Monsell, B. C. (2007). The x-13a-s seasonal adjustment program. *Proceedings of the 2007 Federal Committee On Statistical Methodology Research Conference. URL http://www. fcsm. gov/07papers/Monsell. II-B. pdf*, 515.

Monti, F. (2010). Combining judgment and models. *Journal of Money, Credit and Banking*, *42*(8), 1641–1662.

Robertson, J. C., Tallman, E. W., & Whiteman, C. H. (2005). Forecasting using relative entropy. *Journal of Money, Credit, and Banking*, *37*(3), 383–401.

Rossi, B., & Sekhposyan, T. (2019). Alternative tests for correct specification of conditional predictive densities. *Journal of Econometrics*, *208*(2), 638–657.

Sims, C. A. (2003). Implications of rational inattention. *Journal of Monetary Economics*, *50*(3), 665–690.

Stark, T. (2013). SPF panelists' forecasting methods: A note on the aggregate results of a november 2009 special survey. *Federal Reserve Bank of Philadelphia*.

Svensson, L. E. (2005). Monetary policy with judgment: Forecast targeting. *International Journal of Central Banking*.

Tallman, E. W., & Zaman, S. (2020). Combining survey long-run forecasts and nowcasts with BVAR forecasts using relative entropy. *International Journal of Forecasting*, *36*(2), 373–398.

Waggoner, D. F., & Zha, T. (1999). Conditional forecasts in dynamic multivariate models. *The Review of Economics and Statistics*, *81*(4), 639–651.

Woodford, M. (2013). Macroeconomic analysis without the rational expectations hypothesis. *Annual Review of Economics*, *5*(1), 303–346.

Online Appendix

A Survey Data

Table A.1: Euro area SPF

HICP inflation forecast, y-o-y percent change		
Current calendar year	Annual average	
Next calendar year	Annual average	
Calendar year after next	Annual average	Q3 and Q4 rounds as of 2000, all rounds as of 2013.
One year ahead	Month	
Two years ahead	Month	
Real GDP growth forecast, y-o-y percent change		
Current calendar year	Annual average	
Next calendar year	Annual average	
Calendar year after next	Annual average	Q3 and Q4 rounds as of 2000, all rounds as of 2013.
One year ahead	Quarter	
Two years ahead	Quarter	
Unemployment rate, percent of labour force		
Current calendar year	Annual average	
Next calendar year	Annual average	
Calendar year after next	Annual average	Q3 and Q4 rounds as of 2000, all rounds as of 2013.
One year ahead	Month	
Two years ahead	Month	

Note: Unless otherwise specified, all forecasts are collected as of 1999Q1.

B SPF Variables definitions

We use both fixed horizon and fixed event forecasts from the ECB SPF. Since the definition of each forecasted variable varies with the type of horizon, we outline them in detail below for each variable, namely HICP inflation, real GDP growth and unemployment.

1. HICP inflation

    - Fixed horizon: year-on-year percentage change for the month of reference, quarterly frequency

$$X^{Q}_{t,\text{HICP}} = \left(\frac{Y^{(1)}_{t}}{Y^{(1)}_{t-1}} - 1 \right) \cdot 100 \tag{B.1}$$

where $Y^{(1)}_t = \frac{1}{3}(Y^{(1)}_t + Y^{(2)}_t + Y^{(3)}_t)$, ..., $Y^{(4)}_t = \frac{1}{3}(Y^{(10)}_t + Y^{(11)}_t + Y^{(12)}_t)$.

- Fixed event: year-on-year percentage change for the year of reference, defined as annual average-over-annual average growth rate for the year of reference

$$X^{A}_{t,\text{HICP}} = \left(\frac{\frac{1}{4}\left(Y^{(1)}_t + Y^{(2)}_t + Y^{(3)}_t + Y^{(4)}_t\right)}{\frac{1}{4}\left(Y^{(1)}_{t-1} + Y^{(2)}_{t-1} + Y^{(3)}_{t-1} + Y^{(4)}_{t-1}\right)} - 1 \right) \cdot 100 \tag{B.2}$$

where $Y^{(1)}_t = \frac{1}{3}(Y^{(1)}_t + Y^{(2)}_t + Y^{(3)}_t)$, ..., $Y^{(4)}_t = \frac{1}{3}(Y^{(10)}_t + Y^{(11)}_t + Y^{(12)}_t)$.

2. Real GDP

- Fixed horizon: year-on-year percentage change for the quarter of reference, quarterly frequency

$$X^{Q}_{t,\text{GDP}} = \left(\frac{Y^{q}_{t}}{Y^{q}_{t-1}} - 1 \right) \cdot 100 \tag{B.3}$$

- Fixed event: year-on-year percentage change for the year of reference, defined as annual average-over-annual average growth rate for the year of reference

$$X^{A}_{t,\text{GDP}} = \left(\frac{Y^{(1)}_t + Y^{(2)}_t + Y^{(3)}_t + Y^{(4)}_t}{Y^{(1)}_{t-1} + Y^{(2)}_{t-1} + Y^{(3)}_{t-1} + Y^{(4)}_{t-1}} - 1 \right) \cdot 100 \tag{B.4}$$

3. Unemployment

- Fixed horizon: level as a percentage of the labour force for the month of reference, quarterly frequency

$$X^{Q}_{t,\text{UNE}} = Y_m \tag{B.5}$$

- Fixed event: annual average of the level for the year of reference

$$X^A_{t,\text{UNE}} = \frac{1}{12}\sum_{i=1}^{12} Y_t^{(i)} \tag{B.6}$$

C Measurement equations

We outline here the transformations needed to go from the observed variables to the model implied variables. We use these transformations when constructing our measurement matrix for the measurement equation of the state space model.

C.1 Quarterly y-o-y forecasts

For real GDP and HICP inflation we observe:

$$X^Q_{t,\text{OBS}} = \left(\frac{Y_t^{(1)}}{Y_{t-1}^{(1)}} - 1\right)\cdot 100 \tag{C.1}$$

Where the subscript indicates the year and the superscript indicates the quarter. Simplifying and taking logs of both sides:

$$\log\left(\frac{X^Q_{t,\text{OBS}}}{100} + 1\right) \simeq \log\left(Y_t^{(1)}\right) - \log\left(Y_{t-1}^{(1)}\right) \tag{C.2}$$

Adding and subtracting:

$$\begin{aligned}\log\left(\frac{X^Q_{t,\text{OBS}}}{100} + 1\right) = \log\left(Y_t^{(1)}\right) - \log\left(Y_{t-1}^{(4)}\right) + \log\left(Y_{t-1}^{(4)}\right) - \log\left(Y_{t-1}^{(3)}\right) \\ + \log\left(Y_{t-1}^{(3)}\right) - \log\left(Y_{t-1}^{(2)}\right) + \log\left(Y_{t-1}^{(2)}\right) - \log\left(Y_{t-1}^{(1)}\right)\end{aligned} \tag{C.3}$$

Finally, simplifying:

$$\log\left(\frac{X^Q_{t,\text{OBS}}}{100} + 1\right) = \log\left(\frac{Y_t^{(1)}}{Y_{t-1}^{(1)}}\right) + \log\left(\frac{Y_{t-1}^{(4)}}{Y_{t-1}^{(3)}}\right) + \log\left(\frac{Y_{t-1}^{(3)}}{Y_{t-1}^{(2)}}\right) + \log\left(\frac{Y_{t-1}^{(2)}}{Y_{t-1}^{(1)}}\right) \tag{C.4}$$

For unemployment quarterly forecasts, the measurement equation is trivial, since levels are used both for the observed data and for the state variable. We average monthly unemployment over the quarter to tranform it into quarterly frequency.

C.2 Annual averages y-o-y forecasts

For real GDP and HICP inflation, we observe:

$$X_{t,\text{OBS}}^{A} = \left[\frac{\frac{1}{4}\left(Y_t^{(1)} + Y_t^{(2)} + Y_t^{(3)} + Y_t^{(4)}\right)}{\frac{1}{4}\left(Y_{t-1}^{(1)} + Y_{t-1}^{(2)} + Y_{t-1}^{(3)} + Y_{t-1}^{(4)}\right)} - 1 \right] \cdot 100 \tag{C.5}$$

Simplifying and taking logs of both sides, and using continuous compounding and geometric averaging:

$$\log\left(\frac{X_{t,\text{OBS}}^{A}}{100} + 1\right) = \log\left[\left(Y_t^{(1)}Y_t^{(2)}Y_t^{(3)}Y_t^{(4)}\right)^{\frac{1}{4}}\right] - \log\left[\left(Y_{t-1}^{(1)}Y_{t-1}^{(2)}Y_{t-1}^{(3)}Y_{t-1}^{(4)}\right)^{\frac{1}{4}}\right] \tag{C.6}$$

Finally, adding and subtracting:

$$\begin{aligned} \log\left(\frac{X_{t,\text{OBS}}^{A}}{100} + 1\right) = \frac{1}{4}\Bigg[& \log\left(\frac{Y_t^{(4)}}{Y_t^{(3)}}\right) + 2\log\left(\frac{Y_t^{(3)}}{Y_t^{(2)}}\right) + 3\log\left(\frac{Y_t^{(2)}}{Y_t^{(1)}}\right) + 4\log\left(\frac{Y_t^{(1)}}{Y_{t-1}^{(4)}}\right) \\ & + 3\log\left(\frac{Y_{t-1}^{(4)}}{Y_{t-1}^{(3)}}\right) + 2\log\left(\frac{Y_{t-1}^{(3)}}{Y_{t-1}^{(2)}}\right) + \log\left(\frac{Y_{t-1}^{(2)}}{Y_{t-1}^{(1)}}\right)\Bigg] \end{aligned} \tag{C.7}$$

For unemployment, we also consider the average over the quarter, therefore the annual forecast is calculated as:

$$X_{t,\text{UNE}}^{A} = \frac{1}{4}\sum_{i=1}^{4} y_t^{(i)} \tag{C.8}$$

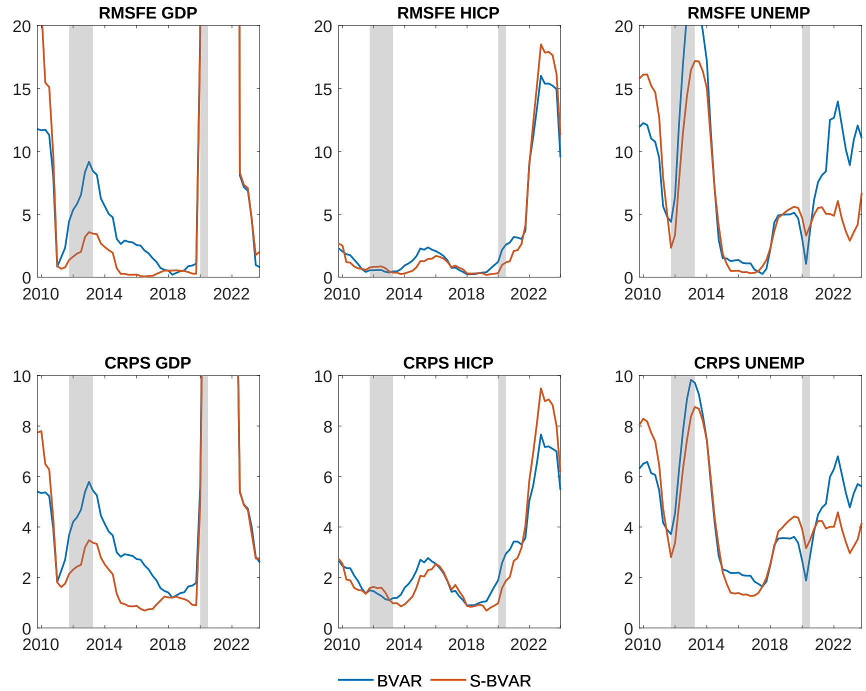
D Example of the ragged edge of fixed event forecasts

Table D.1: HICP Inflation forecast, example for 2024, available forecasts at each quarter

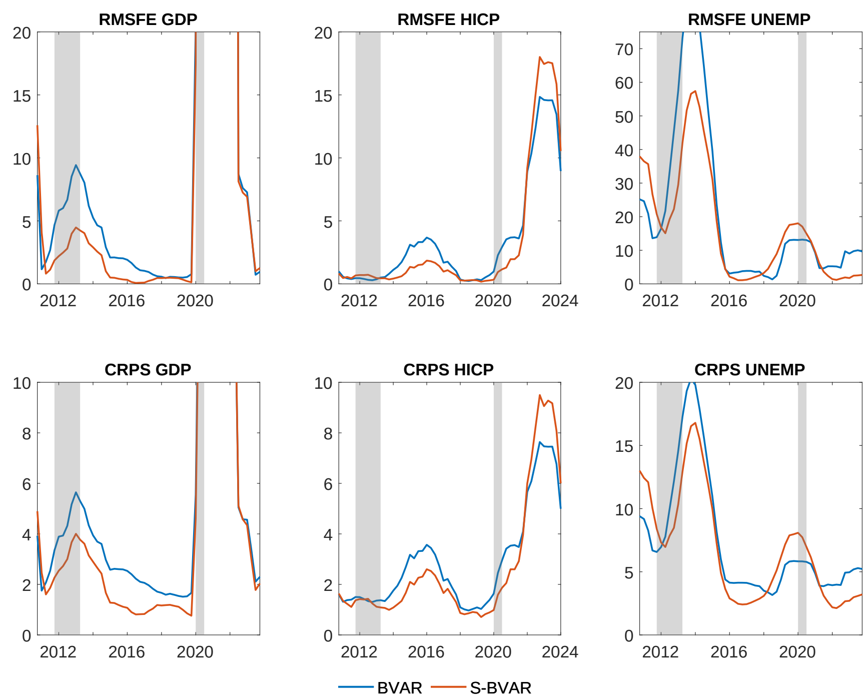
Horizon	2024Q1	2024Q2	2024Q3	2024Q4
FH				
4q	x	x	x	x
8q	x	x	x	x
FE				
1	-	-	-	x
2	-	-	x	-
3	-	x	-	-
4	x	-	-	-
5	-	-	-	x
6	-	-	x	-
7	-	x	-	-
8	x	-	-	-
9	-	-	-	x
10	-	-	x	-
11	-	x	-	-
12	x	-	-	-
19	-	x	-	-
20	x	-	-	-
21	-	-	-	x
22	-	-	x	-

E Additional results

Figure E.1: Rolling absolute scores

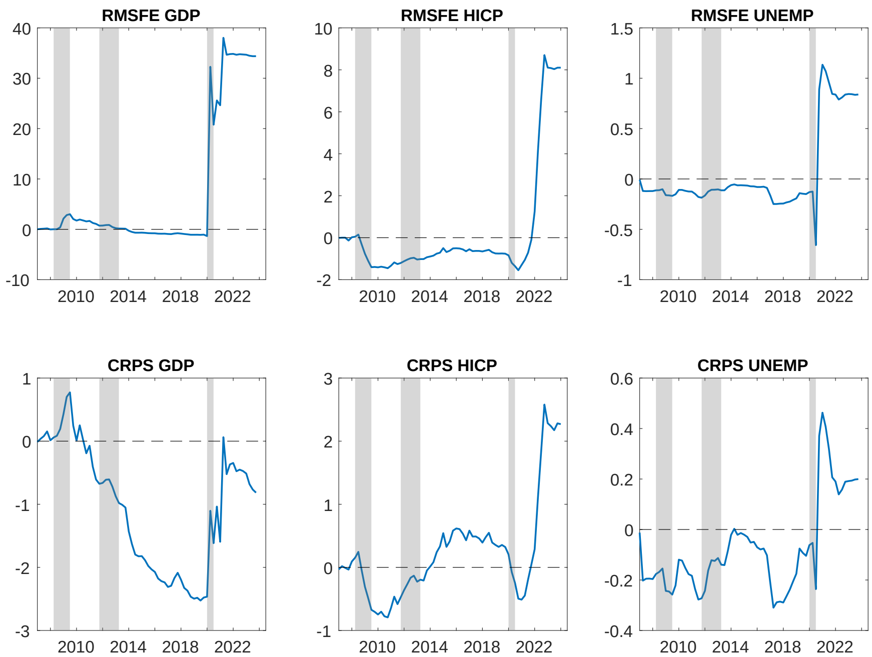


(a) Four-quarter-ahead

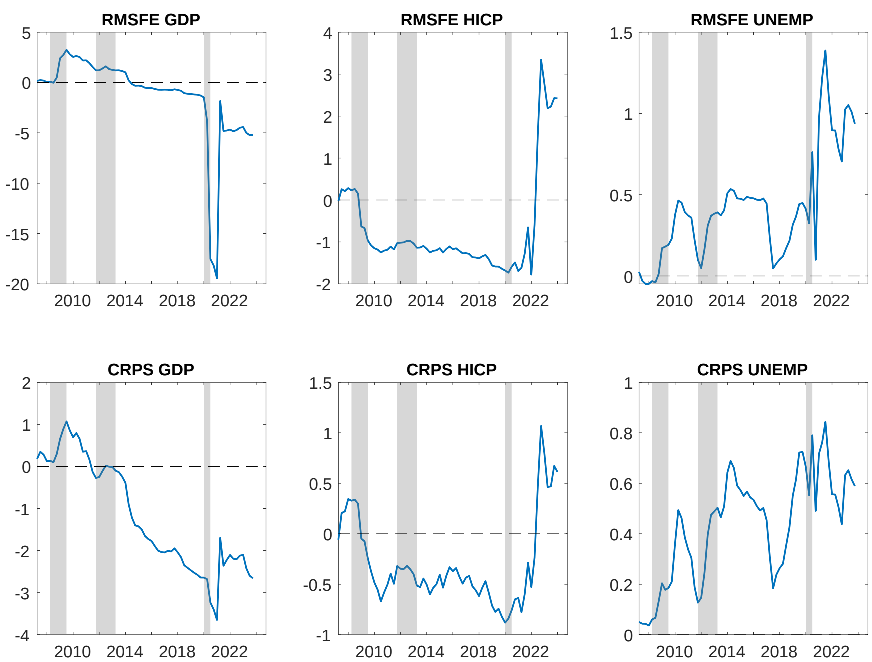


(b) Eight-quarter-ahead

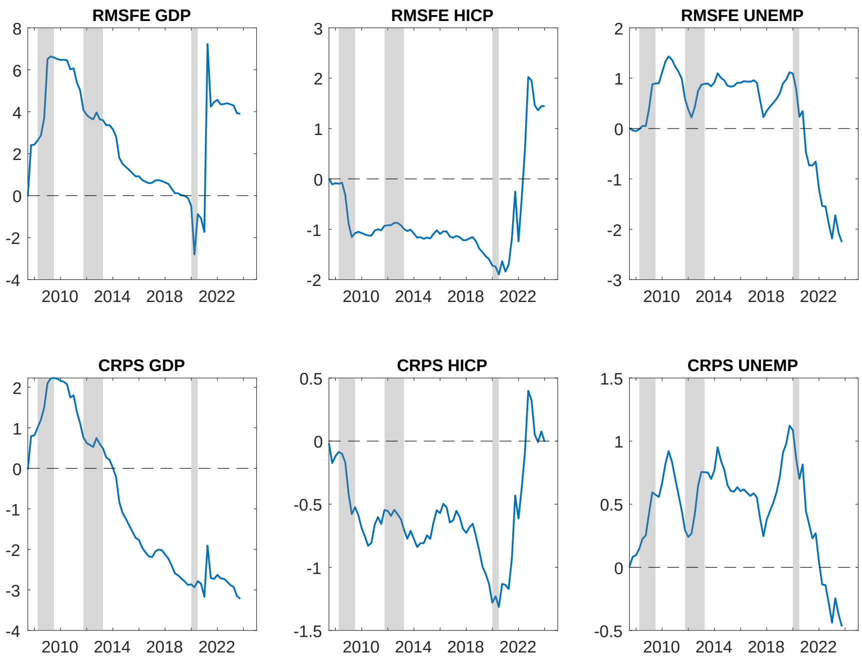
Figure E.2: Cumulative relative scores



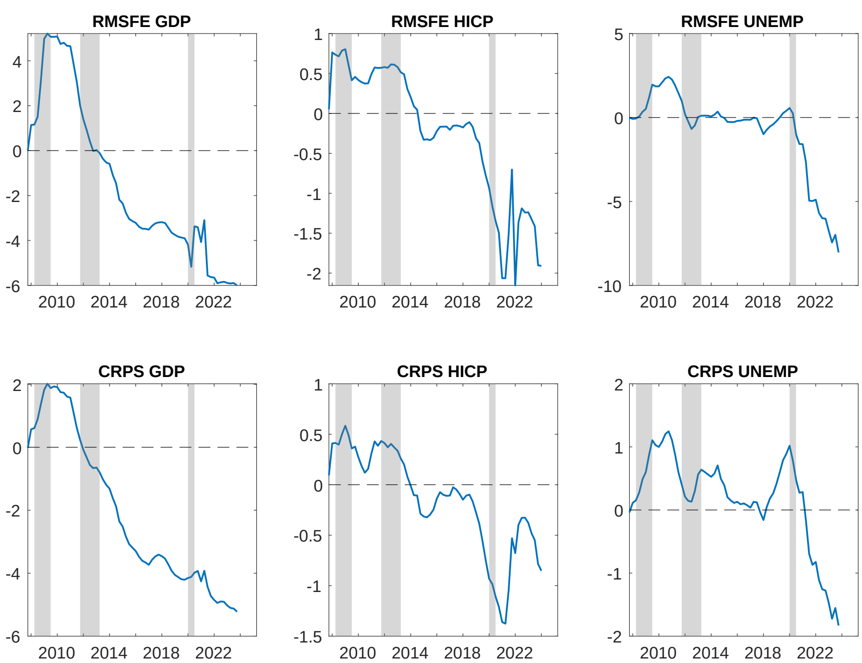
(a) Nowcast



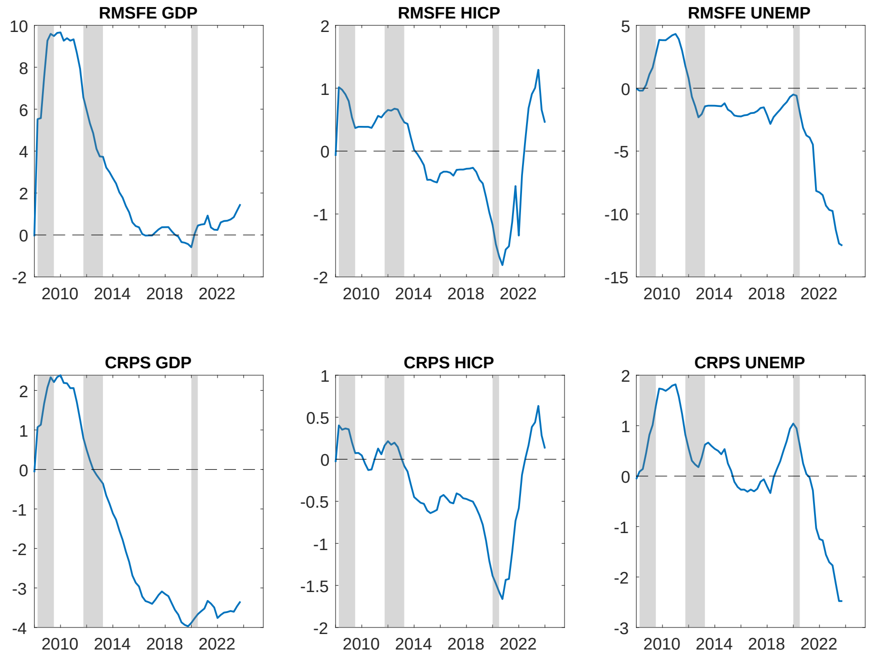
(b) One-quarter-ahead



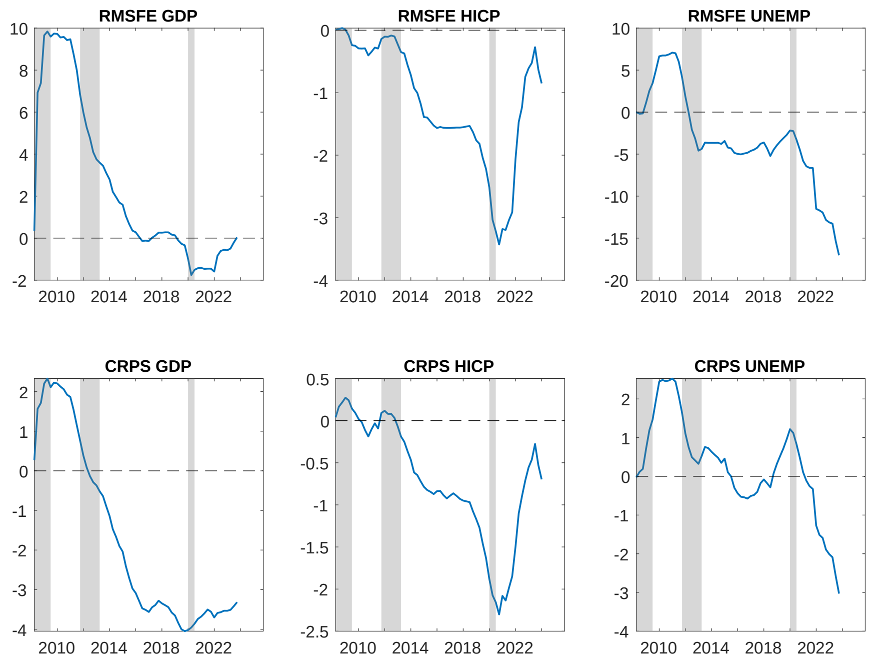
(c) Two-quarter-ahead



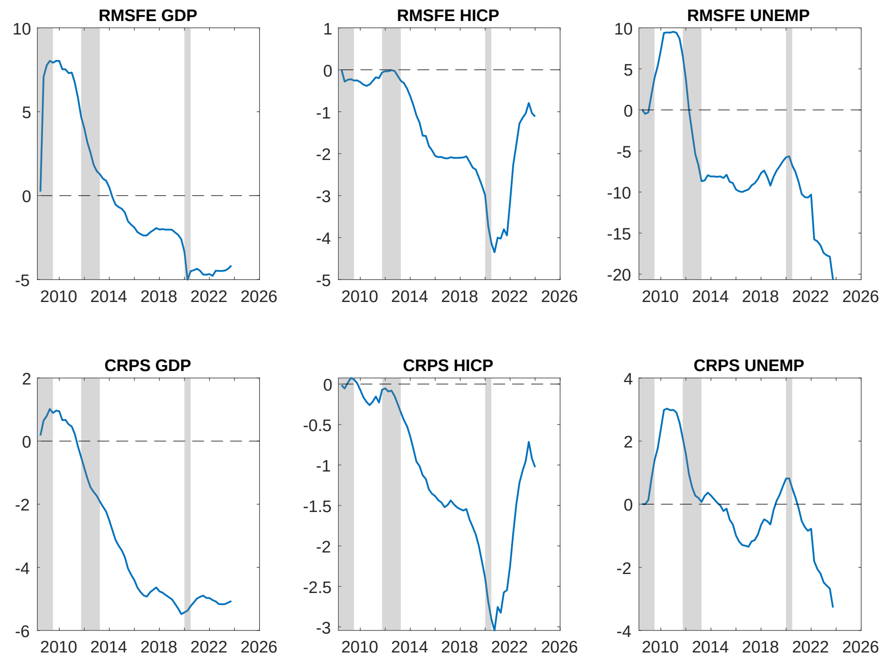
(d) Three-quarter-ahead



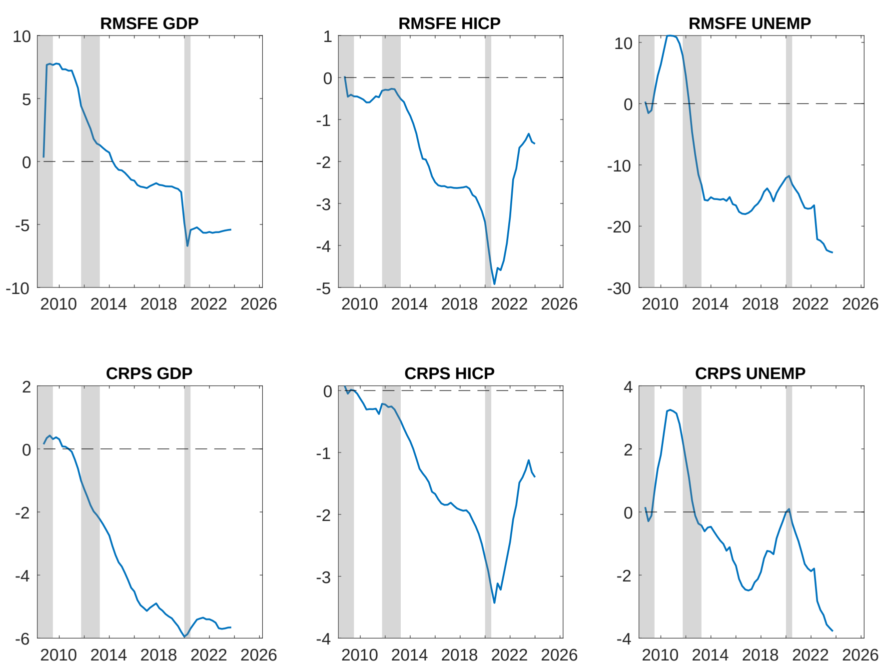
(e) Four-quarter-ahead



(f) Five-quarter-ahead



(g) Six-quarter-ahead



(h) Seven-quarter-ahead

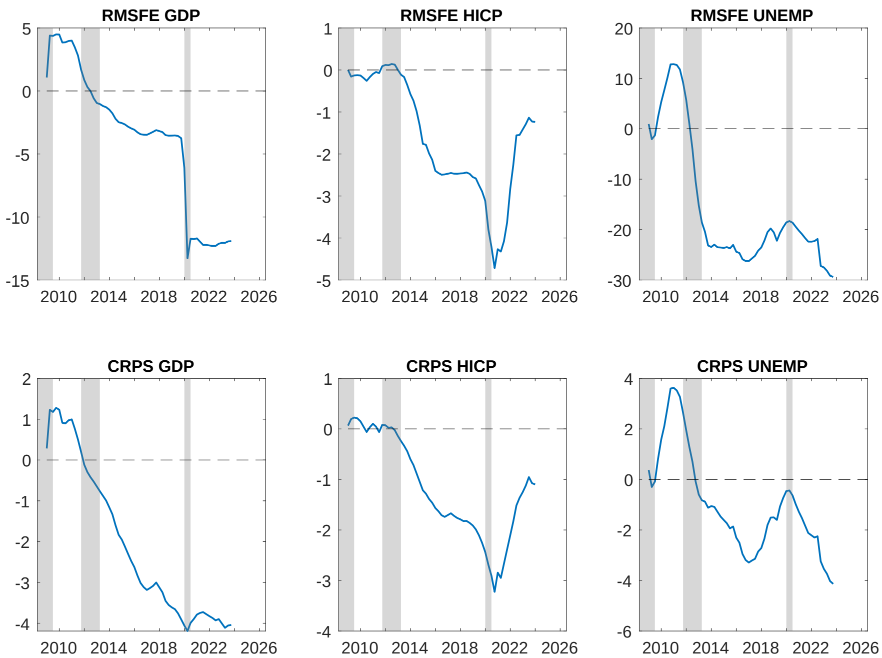


Figure E.2: Eight-quarter-ahead