Combining Bayesian VARs with survey density forecasts: does it pay off?

M. Bańbura[1], F. Brenna[2], J. Paredes[1] and F. Ravazzolo[3]

[1]ECB, [2]KU Leuven, [3]BI Norwegian Business School, Free University of Bozen-Bolzano and RCEA

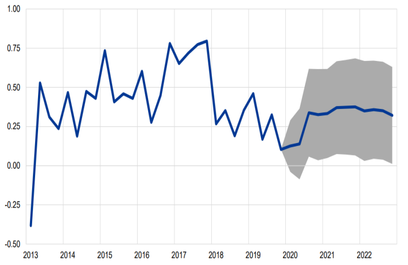
FRBC Conf. - Oct. 06, 2022

Motivation: density forecasts

GDP ranges: pre and post COVID

Euro area real GDP

(quarter-on-quarter percentage changes, seasonally and working day-adjusted quarterly data)



Notes: The ranges shown around the projections are based on the differences between actual outcomes and previous projections carried out over a number of years. The width of the ranges is twice the average absolute value of these differences. The method used for calculating the ranges, involving a correction for exceptional events, is documented in "New procedure for constructing Eurosystem and ECB staff projection ranges", ECB, December 2009, available on the ECB's website.

Euro area real GDP

(quarter-on-quarter percentage changes, seasonally and working day-adjusted quarterly data)

Note: This chart does not show ranges around the projections. This reflects the fact that the standard computation of the ranges (based on historical projection errors) would not, in the present circumstances, provide a reliable indication of the unprecedented uncertainty surrounding the current projections. Instead, in order to better illustrate the current uncertainty, alternative scenarios based on different assumptions regarding the future evolution of the COVID-19 pandemic and the associated containment measures are provided in Box 3.

Source: Eurosystem staff macroeconomic projections for the euro area, March and June 2020.

Motivation: combination and survey information

Tools for forecasting macroeconomic variables:

- From simple random walk to fully-fledged DSGE models;
- Bayesian VARs somewhere in between: **simple** to estimate, **flexible** to adjust, naturally produce **density forecasts** (from posterior densities). These models are now a **standard** tool for central banks forecasts.
- **However**, BVAR models present certain shortcomings:
  - Within time-series world, **large variation across models** in terms of model settings, types of priors, use of off-model information... no "best" specification.
  - Time-series world purely **backward-looking**, hard to get the forecast right in times of heightened uncertainty or "never-happened-before" events.

Our paper

How do we hedge against model uncertainty and adjust for the absence of forward-looking information?

1. We estimate in real-time a wide range of BVAR model specifications, then **optimally combine** the forecast densities obtained.
2. We incorporate **forward-looking survey information**, and analyse to which extent does it improve or hinder forecasts performance.

We find:

1. **Optimally combining several models** improves overall point and density **accuracy**, as well as forecast **calibration**.
2. **Including survey forecasts** on the target's **mean** helps, while on the variance hinders, overall **performance** (accuracy+calibration).
3. Ex-ante tilting performs well $\longrightarrow$ scope for **improving** ex-ante models and **then combine**.
4. COVID case study: SPF information particularly useful in times of unprecedented and fast moving events;

Related literature

- Large literature on Bayesian VARs (see next slide);
- Forecast combination methods: Bassetti et al. (2020); Hall and Mitchell (2007); Jore et al. (2010); Geweke and Amisano (2011).
- Judgement and model forecasts: Galvao et al. (2021) find similar results from tilting forecast combinations to survey moments for the UK; Clements (2018) and Clements (2014) analyse US SPF's density performance; Krüger et al. (2017); Tallman and Zaman (2020); Ganics and Odendahl (2021) apply tilting to a class of individual models; Amisano and Geweke (2017) focus on the combination of several macroeconomic models.

Individual BVARs included in the combination

BVAR model types:

- Minnesota priors with SV (Sims and Zha, 1998; Bańbura et al., 2010)
- Democratic priors with SV (Villani, 2009; Wright, 2013)
- (Survey) Local Mean with SV (Bańbura and van Vlodrop, 2018)
- TVP-SV (Primiceri, 2005)
- UCSV à la Stock and Watson (2007)

More models can be added...

Estimation and forecast

- Target variables: y-o-y growth rate of **euro area HICP** inflation and **GDP**, evaluated at 1- and 2-year ahead horizon (same horizons as for EA SPF).
- Different data set compositions: 3 or 19 variables, aggregated euro area or by country ("big 4": DE, FR, ES, IT).
- **Real-time** recursive estimation, with vintages corresponding to SPF's cutoff dates, and forecasts evaluated from 2000:Q1 to 2019:Q4 (2021:Q3 in the COVID case study).

Optimal linear predictive pool

The model weights are picked so as to maximise the combined predictive likelihood (i.e. log score):

$$\sum_{s=T_1}^{t} log\left(p(y_s; Y_{s-h}, \ldots, Y_1, M)\right), \tag{1}$$

where $p(y_s; Y_{s-h}, \ldots, Y_1, M)$ is the predictive density from model M for y_s given the data $Y_1, \ldots, Y_{s-h}$.

$$w^*_{t+h|t} = \arg\max_{w_i} \sum_{s=T_1}^{t} log\left[\sum_{i=1}^{I} w_i p(y_s; Y_{s-h}, \ldots, Y_1, M_i)\right] \tag{2}$$

where I is the number of models and $w^*_{t+h|t,i}$ is the time-varying weight for model M_i. The weights are constrained to be non-negative and sum to one.

Adding forward looking information: euro area SPF

- Since 1999, asks a panel of experts their forecasts for EA GDP growth, inflation and unemployment, for 1-, 2-, and 5-year horizons;
- Experts need to provide both a point forecast and probabilities for the forecasts to fall within pre-determined ranges;
- The resulting individual responses are aggregated using simple averages;
- We **build a continuous distribution** from the bins, and simulate draws from this distribution, which we use as an additional model in the pool.
- We use both the **point** forecast and the **standard deviation** from the aggregated histograms for tilting models and combination;

Combining survey and model information

Three ways to combine the two sources of information:

1. Include SPF simulated draws as an **additional model** in the optimal pooling;
2. **Tilting "ex-ante"**: use entropic tilting to re-weigh individual model densities so that they take the first and second moments from the SPF, then perform optimal pooling;
3. **Tilting "ex-post"**: use entropic tilting to re-weigh the combined density obtained from optimal pooling.

Combination performance - GDP

	Optimal Pool: *abs. scores*	SPF	Opt. Pool w/SPF	μ tilted ex-ante	μ tilted ex-post	μ & σ tilted ex-ante	μ & σ tilted ex-post
4-q							
CRPS	0.808	**0.994**	**0.997**	**0.935**	**0.932**	**0.966**	**0.971**
LPS	-1.922	-0.627	**0.030**	**0.302**	**0.026**	-0.406	-0.485
PITs	0.042	0.000	0.016	**0.624**	**0.279**	0.000	0.000
8-q							
CRPS	0.994	1.091	1.001	1.080	1.033	1.102	1.099
LPS	-1.973	-1.112	-0.094	-0.042	-0.095	-1.243	-1.303
PITs	0.020	0.000	0.011	**0.099**	0.004	0.000	0.000

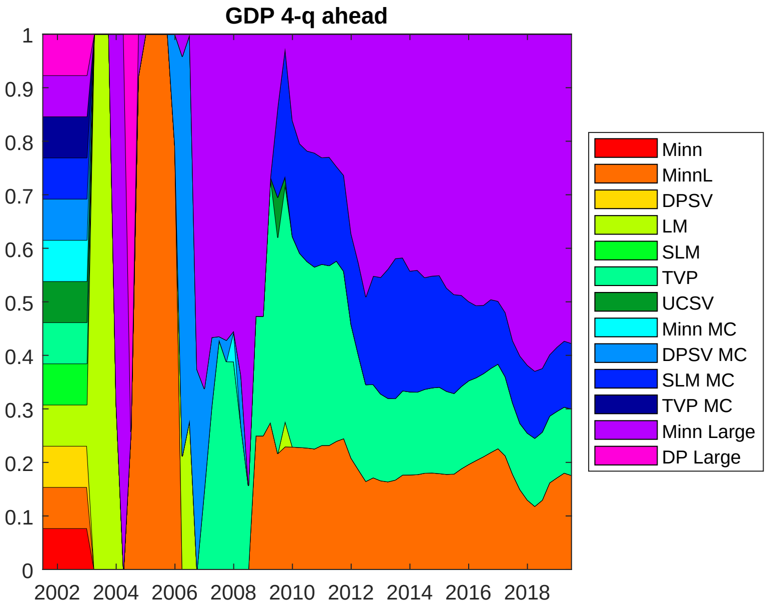
Table: CRPS and LPS: relative accuracy scores with respect to optimal pooling (i.e. first column); PITs: p-values of Berkowitz uniformity test (in absolute terms).

Combination performance - HICP

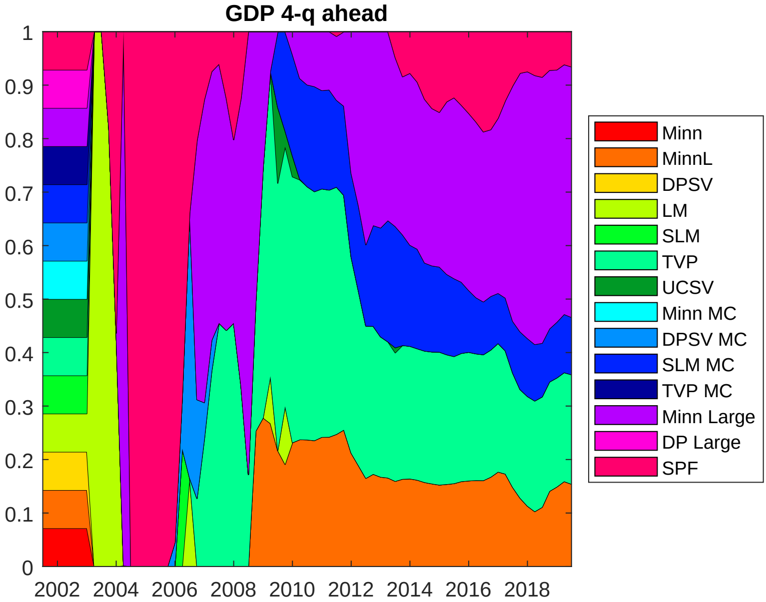
	Optimal Pool: *abs. scores*	SPF	Opt. Pool w/SPF	μ tilted ex-ante	μ tilted ex-post	μ & σ tilted ex-ante	μ & σ tilted ex-post
4-q							
CRPS	0.503	**0.932**	**0.991**	**0.917**	**0.937**	**0.943**	**0.944**
LPS	-1.306	-0.024	**0.003**	**0.117**	**0.056**	-0.007	-0.082
PITs	**0.839**	0.002	**0.704**	**0.218**	**0.156**	0.000	0.000
8-q							
CRPS	0.567	**0.949**	1.020	**0.922**	**0.941**	**0.964**	**0.963**
LPS	-1.429	-0.040	-0.001	**0.082**	**0.032**	-0.263	-0.284
PITs	**0.552**	0.000	**0.961**	**0.368**	**0.232**	0.000	0.000

Table: CRPS and LPS: relative accuracy scores with respect to optimal pooling (i.e. first column); PITs: p-values of Berkowitz uniformity test (in absolute terms).

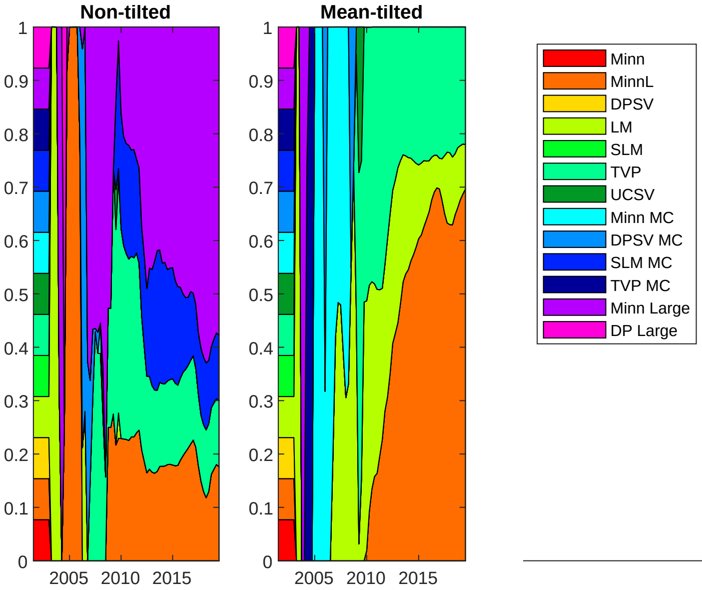
Optimal weights: GDP 1 year ahead



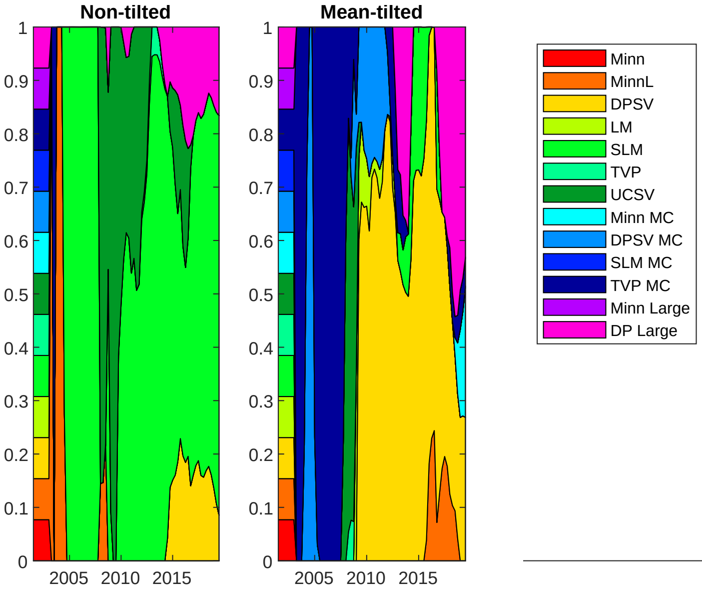
Optimal weights with SPF: GDP 1 year ahead



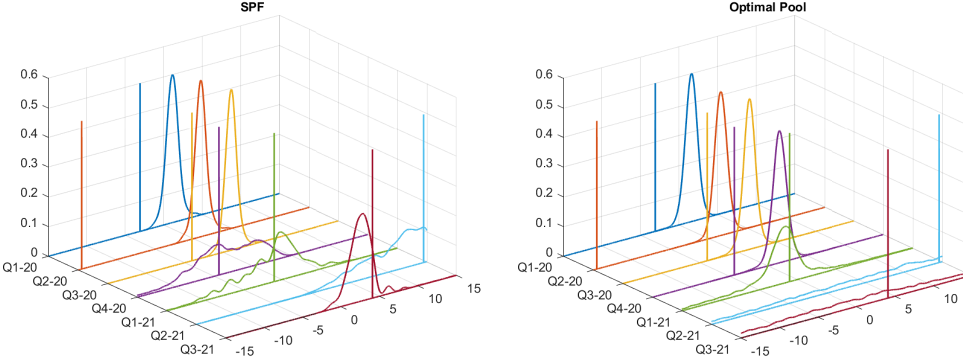
Non-tilted vs tilted weights: GDP 1 year ahead



Non-tilted vs tilted weights: HICP 1 year ahead

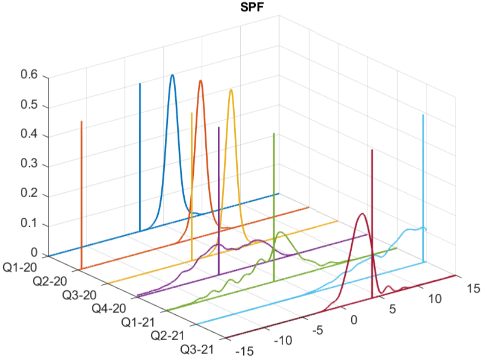


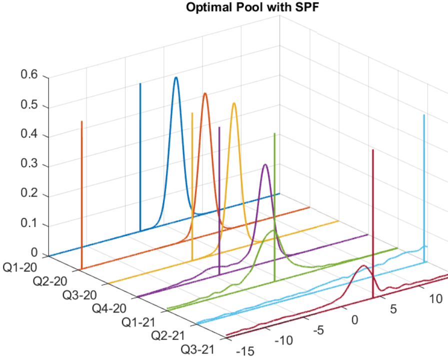
COVID times: a case study



Forecast densities for 2020q1-2021q3 euro area GDP y-o-y growth. The dotted lines are the realizations.

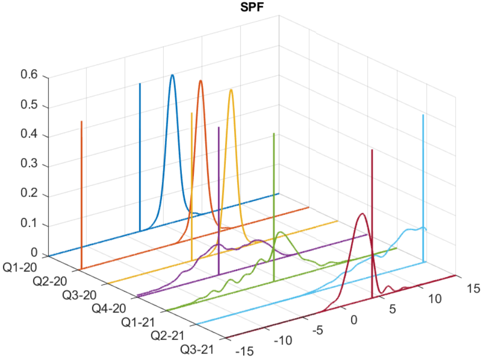
COVID times: a case study

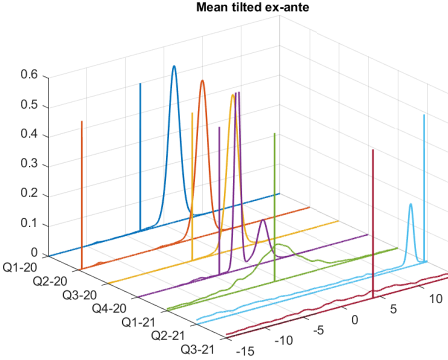




Forecast densities for 2020q1-2021q3 euro area GDP y-o-y growth. The dotted lines are the realizations.

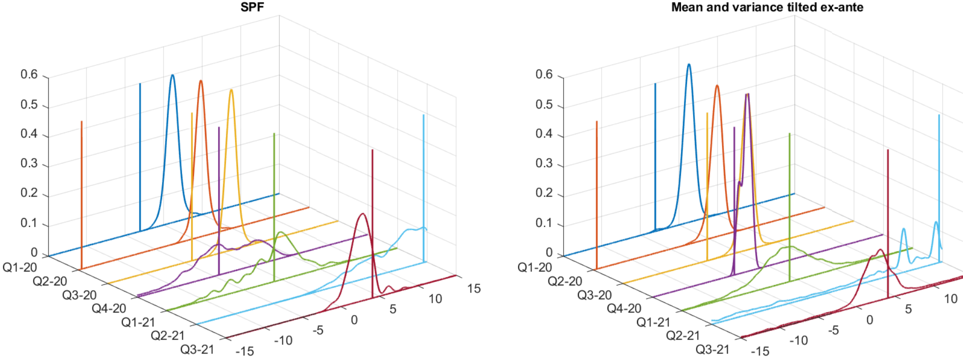
COVID times: a case study





Forecast densities for 2020q1-2021q3 euro area GDP y-o-y growth. The dotted lines are the realizations.

COVID times: a case study



Forecast densities for 2020q1-2021q3 euro area GDP y-o-y growth. The dotted lines are the realizations.

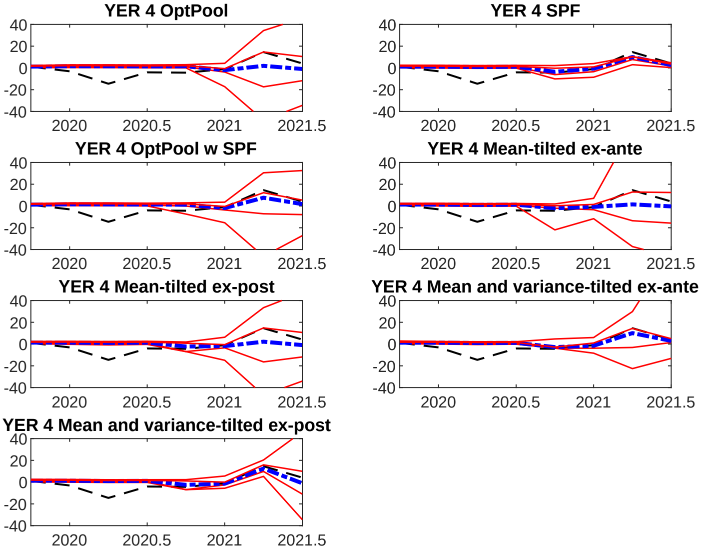
Conclusions

- We evaluate in real time **density forecasts** from a broad range of Bayesian VARs for euro area GDP and inflation; we look at point and density performance over the sample average and at overall calibration, both of individual models and of combinations; we include **judgement** by adding SPF forecasts to the pool of models, or by **tilting** them to SPF **first and second moments**.
- We find large **gains from optimal pooling** for both HICP and GDP.
- The best performing combination (pre-COVID) is the **mean-tilting ex-ante**.
- For both variables, including SPF's **second moments worsens** calibration and density forecasts accuracy.
- In times of high uncertainty and extreme data realisations, higher moments may help, but tilting should be used carefully.

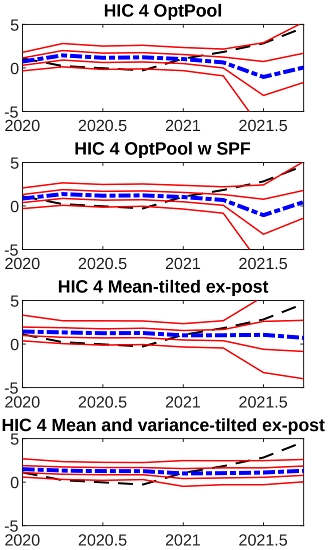
Thank you

Background Slides

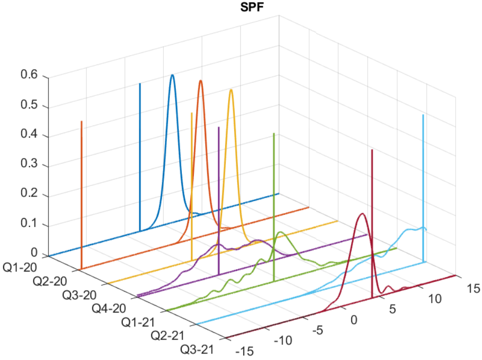
Forecast distributions of one-year-ahead forecasts during COVID - GDP

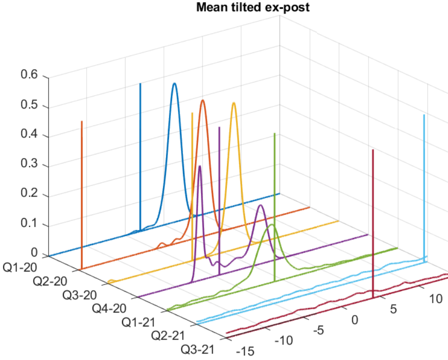


Forecast distributions of one-year-ahead forecasts during COVID - HICP



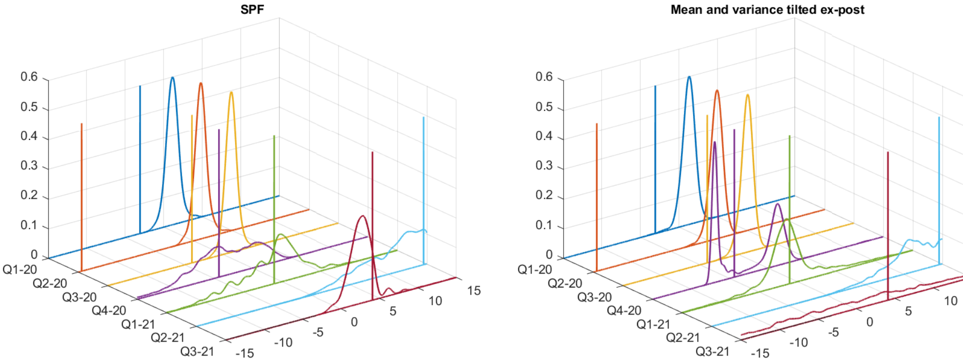
COVID times: a case study





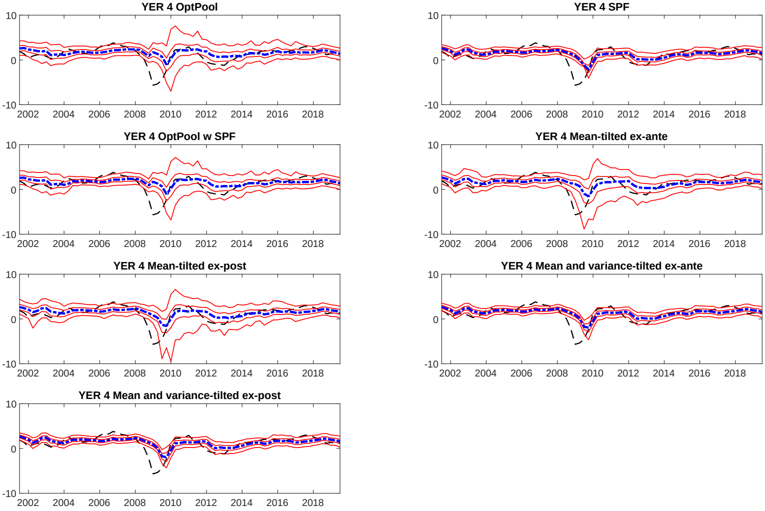
Forecast densities for 2020q1-2021q3 euro area GDP y-o-y growth. The dotted lines are the realizations.

COVID times: a case study

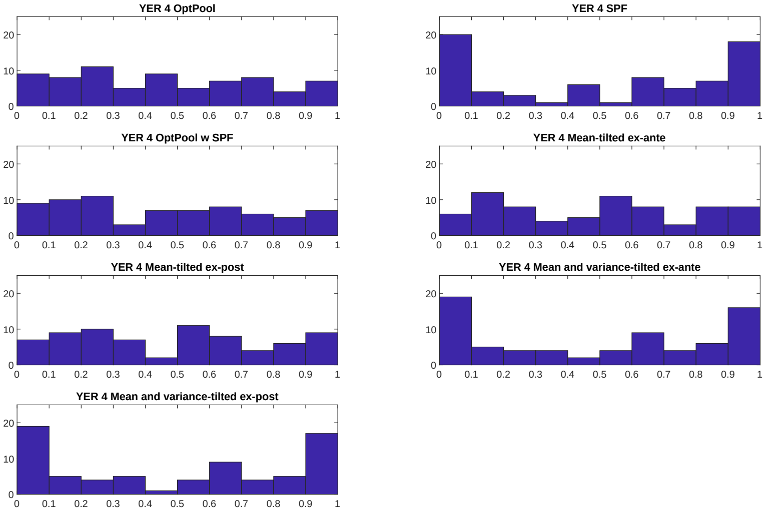


Forecast densities for 2020q1-2021q3 euro area GDP y-o-y growth. The dotted lines are the realizations.

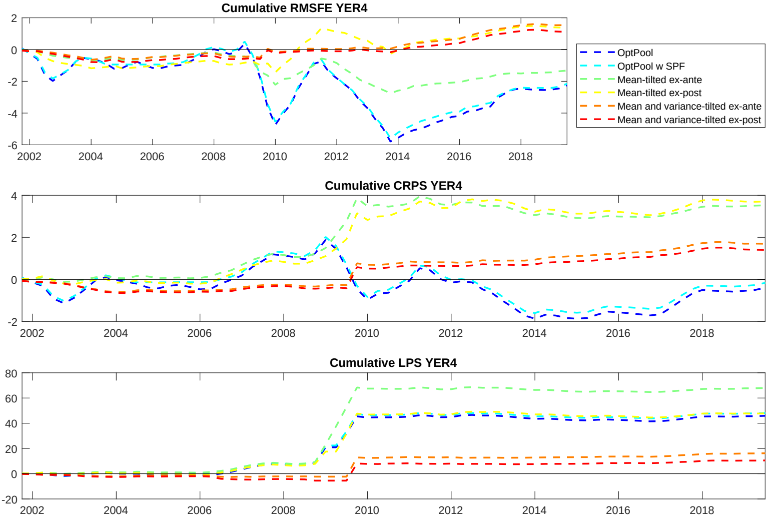
Forecast distributions of one-year-ahead forecasts - GDP



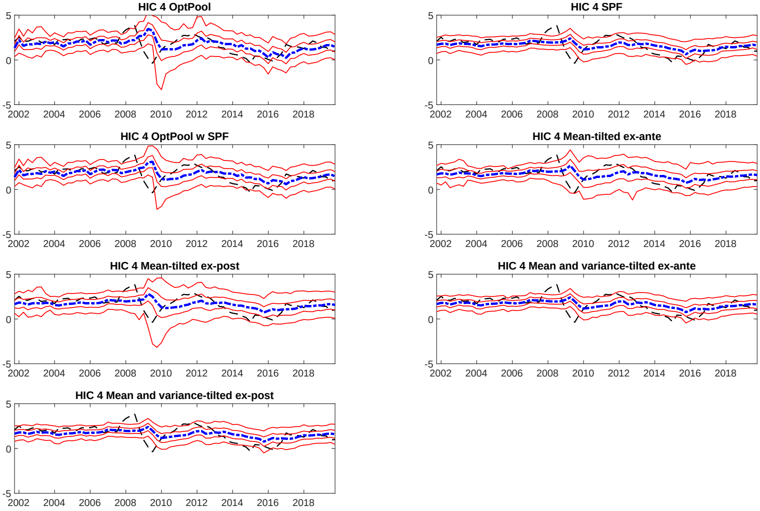
PITs of one-year-ahead forecasts - GDP



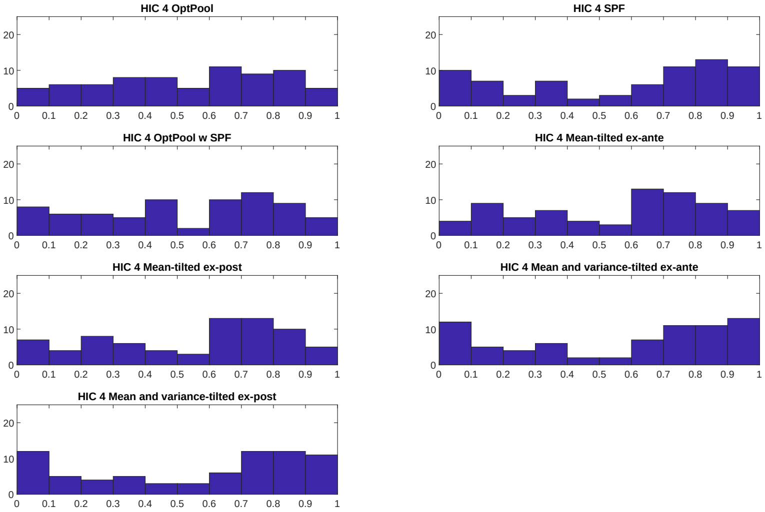
Cumulative relative scores of one-year-ahead forecasts - GDP



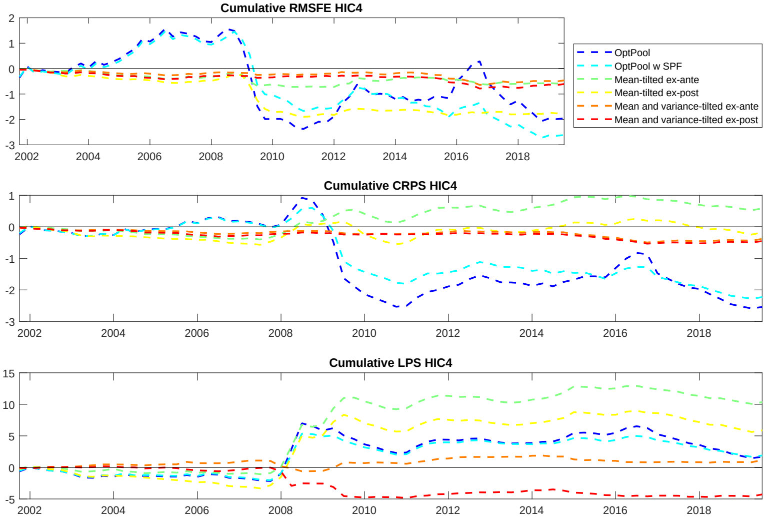
Forecast distributions of one-year-ahead forecasts - HICP



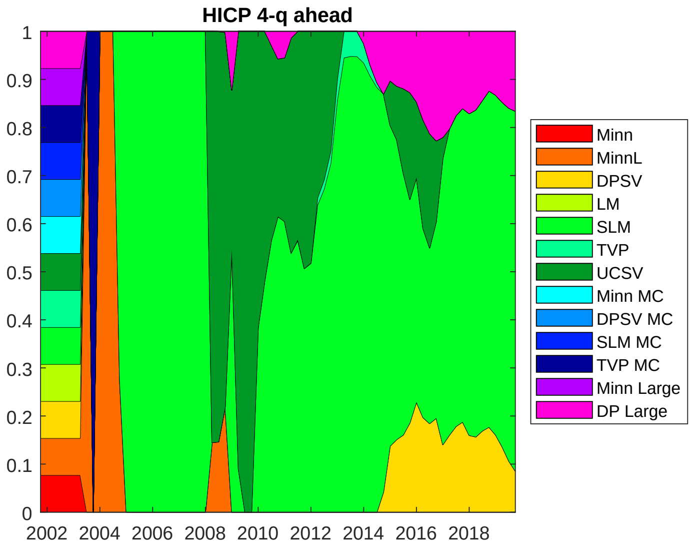
PITs of one-year-ahead forecasts - HICP



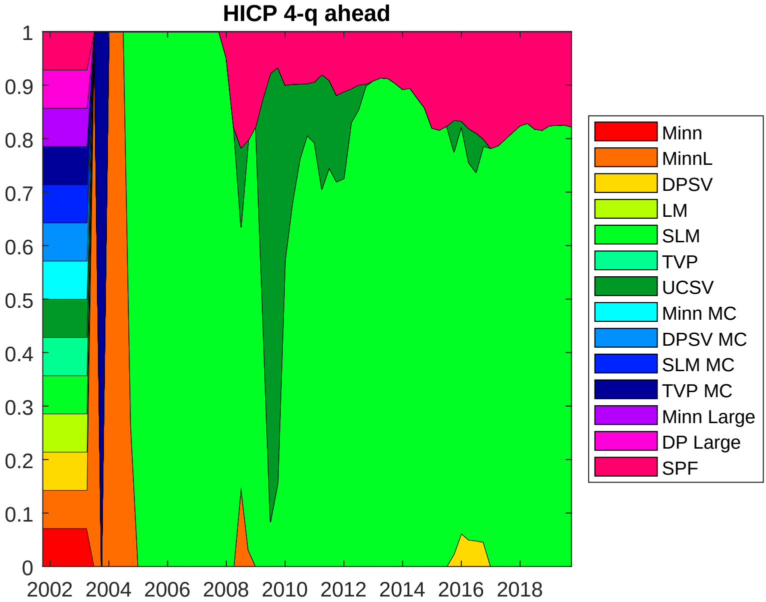
Cumulative relative scores of one-year-ahead forecasts - HICP



Optimal weights: HICP 1 year ahead



Optimal weights including SPF: HICP 1 year ahead



Minnesota priors

Independent normal priors for B_i's. Prior means equal to 0, with the exception of the prior for the diagonal of B_1 (first lag of the dependent variable in each equation) for the specification "in levels", equal to 1. Coefficients for more distant lags are "shrunk" more. The overall degree of shrinkage, as governed by the hyperparameter λ, is set to the standard value of 0.2. The prior for the intercept c is non-informative.

$$y_t = c + \sum_{i=1}^{p} B_i y_{t-i} + \varepsilon_t, \qquad \varepsilon_t \sim N\left(0, H_t\right), \tag{3}$$

Democratic priors

In contrast to the previous version, this parameterisation of the VAR model is in deviation from the unconditional mean μ (sometimes referred to as the "steady state"). Informative normal priors are used for μ. As the mean of the prior we take the long-term forecasts from Consensus Economics. As this type of parameterisation assumes the existence of constant unconditional mean it is only used for variables "in differences". The priors for B_i are the same as above.

$$y_t = \mu + \sum_{i=1}^{p} B_i(y_{t-i} - \mu) + \varepsilon_t, \qquad \varepsilon_t \sim N\left(0, H_t\right), \tag{4}$$

(Survey) Local Mean

VAR in deviation from "local mean", μ_t, that can vary over time as a random walk (reflecting e.g. low frequency changes in demographics, productivity or inflation trend/expectations).

$$y_t - \mu_t = \sum_{i=1}^{p} B_i \left(y_{t-i} - \mu_{t-i} \right) + \varepsilon_t, \qquad \varepsilon_t \sim N\left(0, H_t\right), \tag{5}$$

$$\mu_t = \mu_{t-1} + \eta_t, \qquad \eta_t \sim N(0, V_t) \tag{6}$$

In another version, the local mean is linked to the long-term forecasts from Consensus Economics, z_t:

$$z_t = \mu_t + g_t, \qquad g_t \sim N\left(0, G_t\right), \tag{7}$$

TVP-SV

This is the standard implementation of the VAR where all the coefficients can vary over time, see Primiceri (2005) and Del Negro and Primiceri (2015).

$$y_t = c_t + \sum_{i=1}^{p} B_{i,t} y_{t-i} + \varepsilon_t, \qquad \varepsilon_t \sim N\left(0, \Sigma_t\right), \tag{8}$$

$$c_t = c_{t-1} + \eta_t, \qquad \eta_t \sim N(0, U_t^c), \tag{9}$$

$$B_{i,t} = B_{i,t-1} + \eta_t, \qquad \eta_t \sim N(0, U_t^B), \tag{10}$$

Unobserved Component with Stochastic Volatility

This is a UCSV model à la Stock and Watson (2007) with gamma priors on the error variances in the two stochastic volatility state equations. The model decomposes each variable into a trend and a transitory component, where each component follows an independent stochastic volatility process (see Chan, 2018):

$$y_t = \tau_t + e^{\frac{1}{2}(h_0+\omega_h\tilde{h}_t)}\varepsilon_t^y, \qquad \varepsilon_t^y \sim N(0,1), \qquad (11)$$

$$\tau_t = \tau_{t-1} + e^{\frac{1}{2}(g_0+\omega_g\tilde{g}_t)}\varepsilon_t^\tau, \qquad \varepsilon_t^\tau \sim N(0,1), \qquad (12)$$

$$\tilde{h}_t = \tilde{h}_{t-1} + \varepsilon_t^h, \qquad \varepsilon_t^h \sim N(0,1), \qquad (13)$$

$$\tilde{g}_t = \tilde{g}_{t-1} + \varepsilon_t^g, \qquad \varepsilon_t^g \sim N(0,1), \qquad (14)$$

Measures of relative accuracy: MSFE, CRPS, LPS

1. Mean squared forecast error (*MSFE*):

$$\frac{1}{T_2 - T_1 + 1} \sum_{t=T_1}^{T_2} (y_t - \hat{y}^i_{t|t-h})^2$$

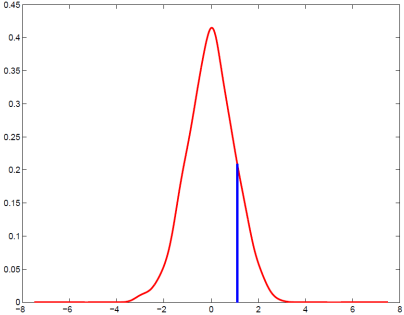
2. Continuous ranked probability score (*CRPS*):

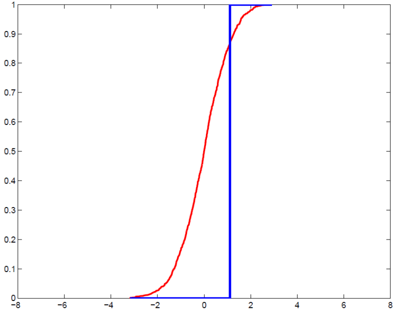
$$\frac{1}{T_2 - T_1 + 1} \sum_{t=T_1}^{T_2} \left(\int_{-\infty}^{\infty} (F(y; y_{t-h}, \ldots, y_1, M_i) - I(y_t \leq y))^2 dy \right)$$

3. Log-predictive score (*LPS*):

$$\frac{1}{T_2 - T_1 + 1} \sum_{t=T_1}^{T_2} log(p(y_t; y_{t-h}, \ldots, y_1, M_i))$$

Examples of LPS and CRPS





Measures of absolute accuracy: Probability Integral Transform (PITs)

$$PIT_t = F(y_t; y_{t-h}, \ldots, y_1, M_i) = \int_{-\infty}^{y_t} f(z_t | I_{t-1}) dz_t, \qquad t = T_1, \ldots, T_2$$

It provides a measure of the model calibration: for well-calibrated predictive distribution (i.e. such that approximates well the actual distribution) the sequence $PIT_{T_1}, \ldots, PIT_{T_2}$ should be uniformly distributed over the interval $[0, 1]$. To test the hypothesis of uniformity, we perform the Berkowitz test. (Berkowitz, 2001)

Entropic tilting

The basic idea of tilting is re-weight a forecast distribution so that it satisfies the moments of interest. The new weights are found so that the new distribution is “close” to the original one, according to the Kullback-Leibler Information Criterion,

$$K(\pi_i^*, \pi_i) = \sum_{i=1}^{k} \pi_i^* log(\pi_i^*/\pi_i)$$

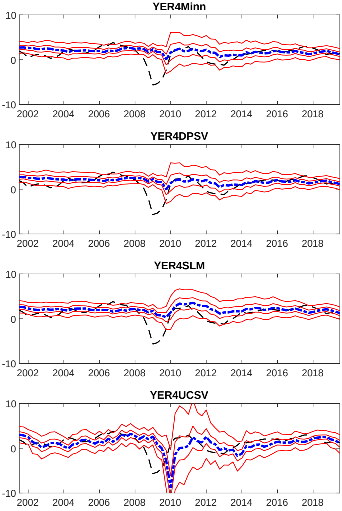
Subject to the following constraints:

$$\pi_i^* \geq 0 \qquad\qquad \sum_{i=1}^{k} \pi_i^* = 1$$

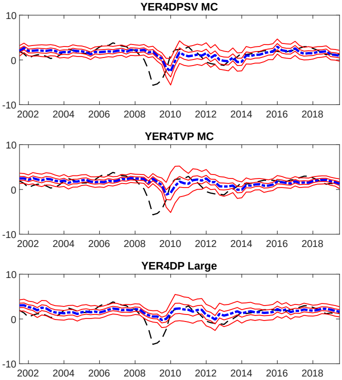
$$\sum_{i=1}^{k} \pi_i^* g(y_i) = \bar{g}$$

The third constraint imposes the moment restrictions and implies that the expectations of a function of the draws from the forecasting distribution should be equal to a fixed quantity.

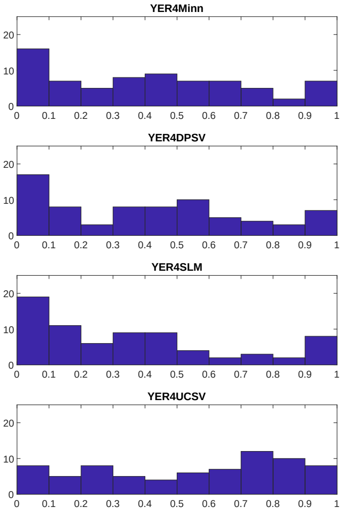
Individual models - GDP 1-year ahead



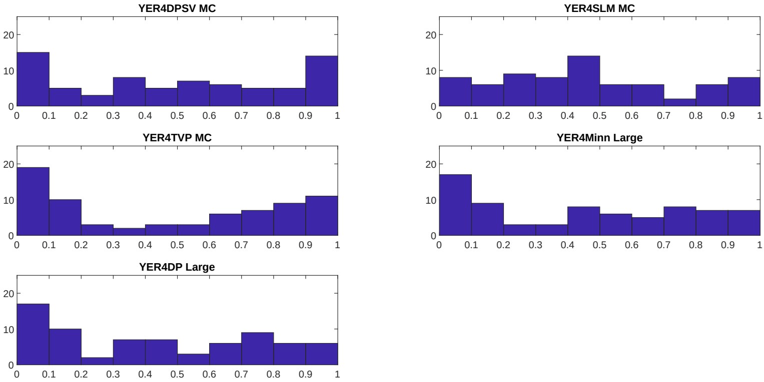
Individual models - GDP 1-year ahead



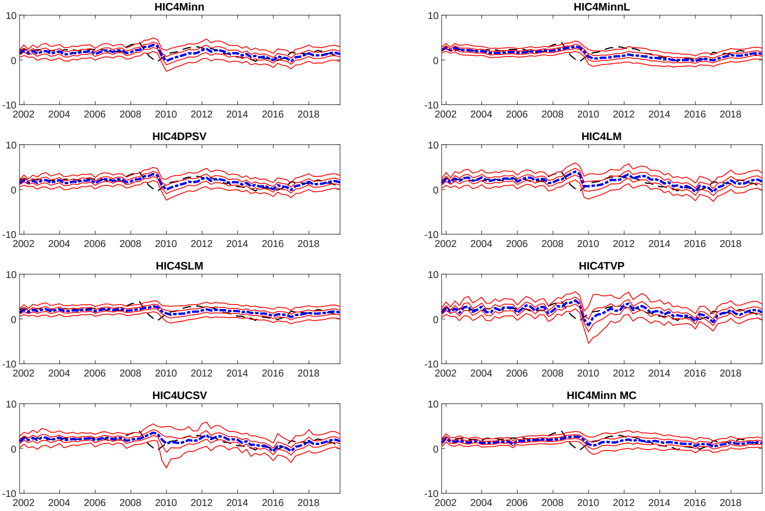
Individual models - GDP 1-year ahead



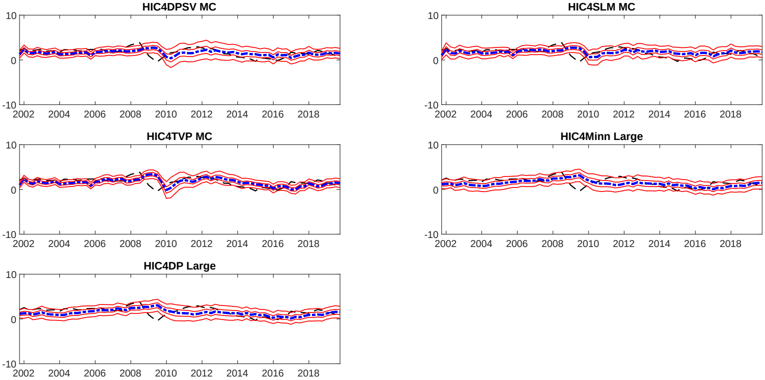
Individual models - GDP 1-year ahead



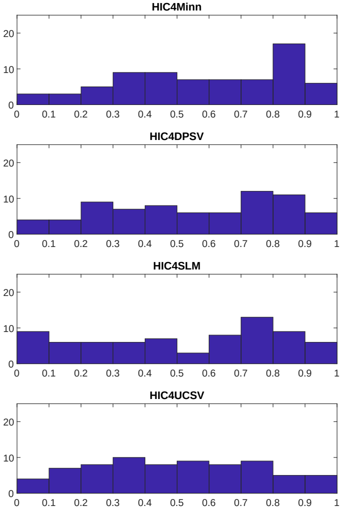
Individual models - HICP 1-year ahead



Individual models - HICP 1-year ahead



Individual models - HICP 1-year ahead



Individual models - HICP 1-year ahead

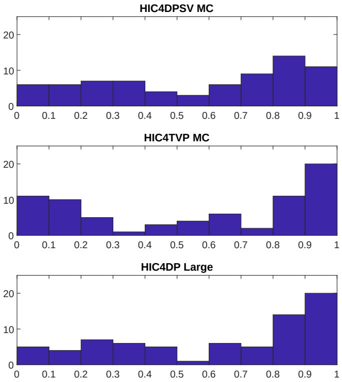


Table: Data set compositions

	Minn Dif	Minn Lev	Dem	LM	SLM	TVP	UCSV
Euro area, 3 variables	x	x	x	x	x	x	x
Euro area, 19 variables	x	-	x	-	-	-	-
Big 4, 3 variables	x	-	x	-	x	x	-

Table: Data set

Variable	Small Model	Medium Model	Transformation
GDP, real	x	x	log-diff
Private consumption, real		x	log-diff
Total investment, real		x	log-diff
Exports XA, real		x	log-diff
Imports XE, real		x	log-diff
GDP deflator		x	log-diff
Total employment		x	log-diff
Short-term interest rate	x	x	levels
Long-term interest rate		x	levels
Lending rate		x	levels
Compensation per employee		x	log-diff
Headline HICP	x	x	log-diff
HICP excluding energy and food		x	log-diff
ESI		x	levels
Foreign demand		x	log-diff
Price of oil in EUR		x	log-diff
Nominal effective exchange rate		x	levels
US short-term interest rate		x	levels
US long-term interest rate		x	levels

References I

Amisano, G. and Geweke, J. (2017). Prediction using several macroeconomic models. *The Review of Economics and Statistics*, 99(5):912–925.

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

Bańbura, M. and van Vlodrop, A. (2018). Forecasting with Bayesian vector autoregressions with time variation in the mean. Tinbergen Institute Discussion Papers 18-025/IV, Tinbergen Institute.

Bassetti, F., Casarin, R., and Ravazzolo, F. (2020). Density forecasting. In *Macroeconomic Forecasting in the Era of Big Data*, pages 465–494. Springer.

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

Chan, J. C. (2018). Specification tests for time-varying parameter models with stochastic volatility. *Econometric Reviews*, 37(8):807–823.

Clements, M. P. (2014). Forecast uncertainty—ex ante and ex post: US inflation and output growth. *Journal of Business & Economic Statistics*, 32(2):206–216.

References II

Clements, M. P. (2018). Are macroeconomic density forecasts informative? *International Journal of Forecasting*, 34(2):181–198.

Del Negro, M. and Primiceri, G. E. (2015). Time varying structural vector autoregressions and monetary policy: a corrigendum. *The Review of Economic Studies*, 82(4):1342–1345.

Galvao, A. B., Garratt, A., and Mitchell, J. (2021). Does judgment improve macroeconomic density forecasts? *International Journal of Forecasting*. Forthcoming.

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

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

Hall, S. G. and Mitchell, J. (2007). Combining density forecasts. *International Journal of Forecasting*, 23(1):1–13.

References III

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

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(3):470–485.

Primiceri, G. E. (2005). Time varying structural vector autoregressions and monetary policy. *The Review of Economic Studies*, 72(3):821–852.

Sims, C. A. and Zha, T. (1998). Bayesian methods for dynamic multivariate models. *International Economic Review*, 39(4):949–68.

Stock, J. H. and Watson, M. W. (2007). Why has US inflation become harder to forecast? *Journal of Money, Credit and banking*, 39(s1):3–33.

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

Villani, M. (2009). Steady-state priors for vector autoregressions. *Journal of Applied Econometrics*, 24(4):630–650.

References IV

Wright, J. H. (2013). Evaluating real-time VAR forecasts with an informative democratic prior. *Journal of Applied Econometrics*, 28(5):762–776.